Identification of genes involved in *Citif1* promoter activation and endoderm differentiation

Thesis

UNRESTRICTED

IDENTIFICATION OF GENES INVOLVED IN *Citif1* PROMOTER ACTIVATION AND ENDODERM DIFFERENTIATION

GABRIELLA LANIA

LAUREA IN SCIENZE BIOLOGICHE
UNIVERSITA' DEGLI STUDI DI NAPOLI"FEDERIDO II"
ITALIA

DOCTOR OF PHILOSOPHY

SPONSORING ESTABLISHMENT
STAZIONE ZOOLOGICA "ANTON DOHRN"
NAPOLI, ITALIA

FEBRUARY 2004

DATE OF SUBMISSION 27 FEBRUARY 2004
DATE OF AWARD 19 MAY 2004

Direct Supervisor	*Prof. Roberto Di Lauro* *SZN - Naples, Italy*
Second internal Supervisor	*Dr. Antonietta Spagnuolo* *SZN - Naples, Italy*
External Supervisor	*Prof. Denis Duboule* *Universitè de Genève Faculté des Sciences Departement De Zoologie et Biologie Animale* *Genève, Switzerland*

Abstract

In tadpole larva of the ascidian *Ciona intestinalis* (one of the most basal chordate) the endoderm is a simple tissue comprising about 500 cells. The commitment of endoderm occurs very early; in a 64-cells stage embryo five blastomeres pairs are restricted to endodermal fate. Although three genes, that exhibit an endoderm specific expression, *Ci*alkaline phosphates, *Cs-lhx3* and *Cititf1* have been cloned, it remains to be elucidated how the first step of endoderm specification take place. As first strategy, I cloned a Ciona SoxE gene; this was found not to be endodermally expressed. In a second strategy, I focused on *Cititf-1*, a gene homologous to mammalian *Nkx2.1* and the earliest zygotic endoderm specific marker isolated from Ciona thus far. Although its temporal and spatial expression seems to be finely regulated, the factors involved in its activation and likely involved in endoderm specification are unknown. The minimal region responsible for enodermal expression of *Cititf1*, has previously been located to a genomic fragment of about 200 base pairs upstream to putative TATA box sequence (Fanelli A. et al., 2003 Developmental Biology. 263, 12-23). Here I describe two imperfect repeats of eleven nucleotides, named R1 and R2, within this region and investigate their involvement in the transcriptional activation of Cititf1 promoter. Biochemical analyses showed that both repeats bind factors from ascidian embryos in a sequence-specific manner. The R2, however, is more important for endoderm specific expression and is able to drive the endodermal expression of a reporter gene. The R1 repeat cooperates with R2 to enhance this expression. A yeast one-hybrid screening was performed, leading to identification of a putative DNA-binding factor that could be involved in the regulation of *Cititf1* in endoderm.

Chapter 1

Introduction

Endoderm differentiation

1.1 General

The development of a multicellular organism from the fertilized egg represents one of the major conquests of evolution. Although developmental processes are different between species, they share some common features. After fertilization, the male and female nuclei fuse and the cleavage process begins. The cleavage consists of a series of mitotic divisions whereby the enormous volume of egg cytoplasm is divided into numerous smaller nucleated cells. These cells are called blastomeres that by the end of cleavage generally from a sphere known as blastula. When the blastula stage is completed, the embryo is divided up into few broad regions that will give rise to the future germ layers: ectoderm, mesoderm and endoderm. At this stage, the future endodermal and mesodermal cells are located outside of the embryo. The blastomeres then undergo dramatic movements wherein they change the position relative to each other. This extensive rearrangement is called gastrulation. This complex process brings the endodermal and mesodermal cells inside of the embryo and the ectodermal cells extend to cover the whole embryo. The ectoderm will form both epidermis and nervous system; the mesoderm will form muscle, cartilage, bone and other internal

organs as heart and kidney; the endoderm, gives rise to the gut associated organs such as lungs, pancreas liver and thyroid. Gastrulation thus involves dramatic changes in the overall structure of the embryo, converting it into a complex three-dimensional structure. During gastrulation, the cells within each of germ layers acquire a specific position, shape and function. Interestingly, in some embryos all this remodelling occurs with little or no increase in cell number or total cell mass (Wolpert L., Principles of development, Oxford University Press, 1998).

Thus, as embryonic development proceeds the complexity of the embryo increases and the fate of each cell becomes more defined. Although the fate of a cell describes what it will normally develop into, it does not imply that cells are restricted to differentiate into a unique type; with time and due to cell-to-cell interaction the developmental potential of each cell becomes more restricted. This commitment or determination process implies a change in gene activity, which starts very early and proceeds gradually during embryonic development.

The identification of genes involved in the molecular pathway responsible for the early determination of the three germ layers, constitutes a major task of developmental biology. A large and detailed amount of studies about ectoderm and mesoderm differentiation has refined the understanding of the molecular mechanisms that generate their cellular diversity. More recently, studies in several animal models, have begun to shed light on the molecular determinants of the endoderm, a germ layer previously less understood.

The bulk of information on endoderm commitment comes from studies in *Xenopus*, zebrafish, *Caenorhabditis elegans* and in sea urchin; in the last few years, studies on Ascidians have also

lend to the identification of genes involved in differentiation of this tissue.

1.2 Definition and derivation of endoderm

The examination of any embryo at the blastula stage gives no indication of how different regions or cells will develop. By following the fate of individual cells, or a group of cells, a fate map on the blastula surface can be constructed that shows or "maps" the larval or adult structure onto the region of the embryo from which it derived. One way of making a fate map is to stain various parts of the surface of an early embryo with a lipophilic dye and observe where the labelled region ends up. Individual cells can also be labelled by injecting high molecular weight molecules such as rhodamine-labeled dextran[1] , which cannot pass through cell membranes and is therefore restricted to the injected cell and its progeny. These fate-mapping studies have revealed that the progenitors of endoderm consist in a group of cells localised in the innermost area of the embryo.

In zebrafish, endoderm progenitors are found at the lateral margin of the blastoderm[2] intermingled among mesoderm progenitors. They are collectively termed mesendoderm (fig.1a). At the early blastula stages these cells are still bipotential and can give rise to mesoderm or endoderm, they become fate restricted at late blastula stage (Kimmel CB et al., 1990). In detail, the endoderm progenitors originate from the four most marginal blastomers tiers, during gastrulation these cells migrate over the extra embryonic yolk sincytial layer

---

[1] *The rhodamine fluoresces red in UV light, so the rhodamine dextran can be easily detected under UV microscope.*

[2] *In fish the blastoderm is defined as the post-cleavage embryo*

(YSL, an important source of inductive signal), involute and occupy the space overlaying the YSL. During gastrulation the mesendoderm cells differentiate and this process is characterised by morphological differences that arise between mesodermal and endodermal cells. The latter become increasingly flatter and lose their polar appearance and obtain filopodial processes. In contrast, neighbouring cells undergo minimal cell-shape change, remaining spherical and lacking filopodia throughout gastrulation. Thus, by late gastrulation the endoderm and the mesoderm are distinguishable tissues (Warga RM et al, 1999).

In amphibian, the endoderm progenitors exist as a more discrete population and originate from yolky cells at the vegetal pole of the blastula (fig. 1b). When vegetal poles are explanted and cultured in isolation up to the mid-blastula stage, they continue to express endodermal markers (Henry GL et al.1996; Hudson C et al., 1997). However, if they are disaggregated and cultured in vitro, the expression of these markers decays by mid-gastrula transition (Clements D et al, 1999; Yasuo H et al., 1999). These evidences indicate that the fate of vegetal pole is pre-programmed to become endoderm precursor cells. However cell-cell interaction is required for maintenance of the definitive endodermal differentiation. Moreover, it has been demonstrated that a single vegetal pole cell from the morula[3] and from the mid-blastula stage can contribute to the progeny of all germ layers when transplanted into the blastocoel[4] of a host. In contrast, cells from the vegetal pole at the early gastrula stage can only contribute to endoderm differentiation (Barker N et al., 2000). Hence, the endoderm progenitor cells acquire a restriction in their fate at the mid-blastula stage but

---

[3] *A morula is the very early stage in a embryo when the cleavage has resulteed in a sphere of cells*
[4] *The blastocoel is the fluid-filled cavity that develops interior of a blastula.*

become determined only by the beginning of gastrulation.

In amphibia, the presumptive dorsal endoderm is also a source of inductive signals, known as the Nieuwkoop centre (Nieuwkoop PD, 1973). These inductive signals generate an equatorial mesodermal compartment midway between the pigmented animal pole and the vegetal pole. Therefore, the mesoderm and the endoderm precursor cells are initially in close anatomical proximity and only separate completely during gastrulation. The failure of a successful endoderm development frequently results in an increase in mesoderm; conversely, the expansion of the endoderm occurs at expense of the mesoderm (Henry GL et al, 1998; Kikuchi Y et al., 2001).

In mammalian embryos the beginning of gastrulation is evidenced by the formation of a structure called the primitive streak. The primitive streak is found at the posterior of the epiblast[5] that elongates towards the future anterior end of the embryo. The primitive streak is necessary for gastrulation to occur properly and also seems to be involved in cell fate specification of mesendoderm progenitor cells. The fate map of the mouse embryo has shown that the most definitive endoderm cells originate from the anterior part of primitive streak (fig.1c). During gastrulation, the mesoendodermal cells accumulate at the anterior end of the primitive streak, involute through the primitive streak, migrate along the midline and give rise to mesoderm and definitive endoderm. The endoderm precursor cells are thought to intercalate into overlying visceral endoderm layer[6] , eventually displacing the cells from the visceral layer.

---

[5] *The epiblast is a group of undifferentiated cells that give rise to the embryo proper.*
[6] *The visceral endoderm is derived from the primitive endoderm and contributes to extra-embryonic tissues.*

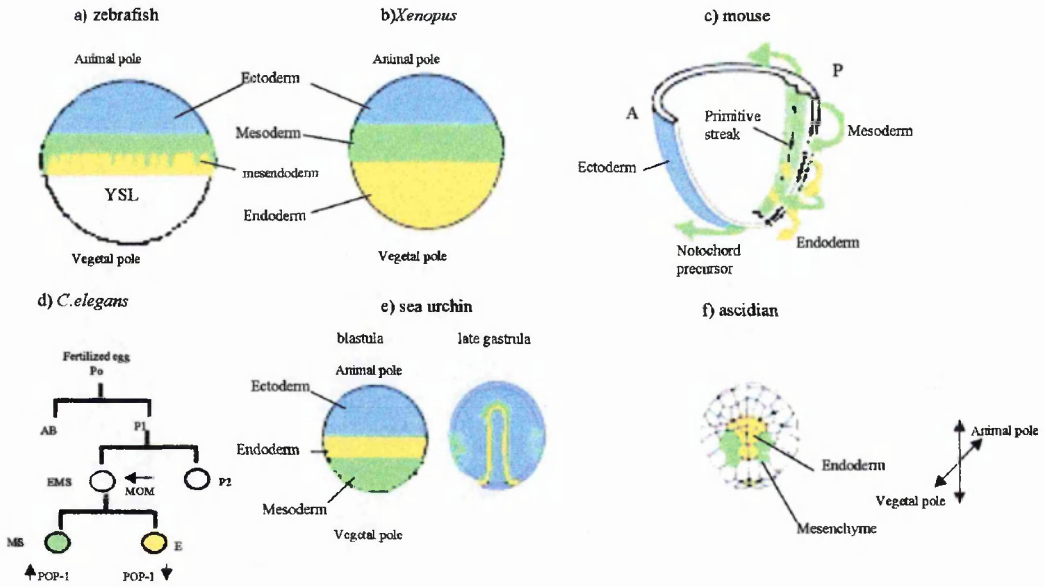


Fig.1: Fate map of endoderm and mesoderm precursor cells in a) fish, b) frog, c) mouse, d) worm, e) sea urchin and f) tunicate.

The anatomic relationship between endoderm and mesoderm, has led to the concept of a mesendodermal field (Kimelman D at al, 2000) that is separated from the ectoderm. One conclusion that arises from a large body of work is that the early events of differentiation are characterised from a initial separation of the mesoendoderm from the ectoderm. Later, the mesendoderm segregates into endoderm or mesoderm depending on the appropriate combination of signalling in each compartment.

The origin of endoderm and mesoderm from the same precursor cells is conserved across metazoan phylogeny. In the nematode *Caenorhabditis elegans*, the endodermal germ layer is established as a single cell, called E, which derives from the division of the mesoendodermal precursor cell (EMS) (fig.1d). The EMS divides into an MS cell (that produces mesodermal muscles) and an E cell (that produces intestinal endoderm). The segregation of mesoderm and endoderm is regulated by a signal that originates from a cell adjacent to the

posterior side of EMS (Goldestain B, 1992) known as the P2 cell. The P2 cell produces a signal that interacts with the EMS cell and instructs the EMS daughter that is next to it to become the E cell. Mutation in genes encoding proteins involved in this pathway, leads to the transformation of E to MS (More of MS, MOM mutant) or the transformation of MS to E (Posterior Pharynx Deleted, POP mutant), (Lin R et al., 1995; Lin R et al., 1998).

A mesoendodermal field can be also identified in sea urchin, where both endoderm and mesoderm derive from the lower part of vegetal pole of the embryo (fig. 1e). The vegetal half of a sea urchin embryo consists of a ring of 8 cells called veg1, another ring of 8 cells called veg2 and two groups of four cells called large and small micromeres at the vegetal pole. All of the endomesoderm derives from these vegetal components. During blastula formation, the veg2 progeny and the micromere lineages form a thickened disk at vegetal end of the embryo forming a vegetal plate. The micromere progeny are located at the centre of this disc. All mesoderm cell types derive from this central region, while the surrounding peripheral region gives rise to all endodermal precursors cells (Ruffins SW et al., 1996). By gastrulation, the mesoendoderm specification has occurred and the mesoderm and endoderm precursors cells moves inside of the embryo from the vegetal pole. The first event is the entry into blastocoel of primary mesenchyme (mesoderm) cells from the centre of vegetal plate. Subsequently, the endoderm together with the secondary mesenchyme starts to invaginate leading to the formation of a primitive gut or archenteron. As the invagination proceeds, the archenteron elongates and finally contacts the basal surface of the presumptive ectoderm at the animal pole. The fusion between the two tissues results in formation of

the mouth opening. Through this process the mouth, gut and anus are formed.

In the lower chordate Ascidian, the developmental fate of the endoderm (but also of the other tissues as epidermis and some muscle cells) is determined at very early stages of embryogenesis due to the presence of pre-localized cytoplasmic factors in the egg (Whittaker JR 1979, Nishida H 1996). Therefore if these cells are lost during development, others cannot replace them. Moreover, as demonstrated by the detection of alkaline phosphates activity in the endodermal tissue, the vegetal blastomeres differentiate in endoderm cells when isolated from an embryo at 8-cell stage, strongly suggesting an autonomous differentiation mechanism in endoderm development (Nishida H, 1992). All endodermal progenitor cells are clearly distinguishable at blastula stage. These cells have been identified by tracer injection (Nishida H, 1987) and their determination occurs at 32-cell stage embryo. Before the gastrula stage, at 110-cell stage, the endoderm precursors cells consist of a group of 10 cells in the vegetal pole of the blastula (fig.1f). These cells will give rise to the endoderm of the trunk in swimming tadpole larva (Nishida H, 1987). After metamorphosis of the larva, in the sessile adult organism, the endoderm of the trunk will differentiate into the gill opening, digestive organ and endostile, which is thought to be homologous to the vertebrate thyroid (Ericson LE et al., 1990).

In the past few years, a large body of work has increased our knowledge in the process of endoderm formation both in vertebrate and invertebrate model systems. In particular, recent studies have provided new insights into the molecular mechanisms responsible for the delineation of the endoderm in Xenopus and in zebrafish and the conserved role of some of

these molecules in mouse development. These studies reveal a high degree of conservation in some of the transcriptional regulators and highlight the divergence in the intracellular signalling events leading to formation of this tissue. For example, transcription factors such as GATA are implicated in endoderm formation across the phyla. On the other hand, signalling molecules such as Nodal appear to be vertebrate specific, while signalling activated by Wnt molecules seems to be specific for endoderm differentiation in invertebrates.

1.3 The hierarchy of signals in endoderm formation

Experimental evidence from studies in vertebrate animal models favours a two-step model in early endoderm differentiation; in an initial phase, maternal and Nodal related-factors factors within the mesoendodermal field promote the endoderm and mesoderm differentiation. In a second phase, downstream effectors generate and stabilize a programme for commitment to endoderm cell lineage. This model comes from several lines of experimental evidence in Xenopus, but a strong conservation in the pathway of endoderm differentiation in zebrafish has also been observed. Additionally, murine counterparts of some of these molecules have also been isolated.

In amphibia, there are currently two classes of candidates for maternal molecules that initiate endoderm development, the transforming growth factor-b family members, such as activin and Vg1, and T-box transcription factor, VegT.

While no expression of activin is detected in the egg, activin-like proteins are present, although to date nothing is known about their function (Asashima M et al., 1991). Conversely, the Xenopus Vg1 gene is detected as a maternal mRNA localized to the prospective endoderm, and mature Vg1 protein can induce both endodermal and mesodermal markers in embryonic cells (Weeks DL et al., 1987). The role of Vg1 in endoderm differentiation has been also demonstrated using a mutant dominant negative form. Blocking the Vg1 signalling prevents the formation of endoderm and mesoderm (Joseph EM et al., 1998). A homologue of Vg 1 has also been identified in zebrafish (Dohrmann CE et al., 1996). The zebrafish zDVR-1 gene, like Xenopus Vg1, is present maternally as an unprocessed precursor protein that is distributed ubiquitously along the future dorsoventral axis. As processed Vg1, mature zDVR-1 is a potent inducer of axial mesoderm, but there is no information on its involvement in endoderm differentiation.

The Xenopus VegT, a T-box containing protein, represents the single maternal transcription factor involved in endoderm differentiation isolated to date. Xenopus embryos depleted of maternal VegT do not express endodermal genes and this expression is rescued upon injection of the mRNA encoding for VegT (Zhang J et al. 1998; Clements D et al, 1999). VegT appears to promote endoderm differentiation by activating TGFb/nodal related factors. In fact, Nodal or the TGFb-like proteins, such as Xnr1, Xnr2, Xnr4 and derriere, are able to rescue the phenotype of VegT depleted amphibian embryos (Xanthos JB et al., 2001). Moreover, the ability of VegT mRNA injection to rescue the VegT depleted embryo, is blocked by coinjection of a non-secreted mutant form of Xnr2 and derrière (Xanthos JB et

al., 2001). These results indicate that only a basal level of activity of early endodermal genes is initiated and maintained by action of maternal VegT. It is possible that VegT regulates endoderm formation by also controlling Nodal expression.

In zebrafish, as well as in Xenopus, the overexpression of Antivin, a competitive inhibitor of Activin/Nodal signalling pathway leads to a complete loss of endoderm (Alexander J et al, 1999; Osada S et al., 1999). Furthermore, the expression of a constitutively active form of the TGF-b receptor is sufficient to induce endodermal gene expression (Alexander J et al., 1999). Several lines of experimental evidence support the importance of nodal signalling in zebrafish. The double mutant Cyclops (cyc) and Squint (sqt), Nodal related proteins, lacks all endoderm (Feldeman B et al., 1998). The same phenomenon was observed in embryos lacking both maternal and zygotic one-eyed pinhead (MZoep), an EGF-CFC family member (Zhang J et al., 1998) believed to be a co-receptor for Nodal molecules (Yeo C et al., 2001). The implication of Nodal signalling, in endoderm differentiation is also demonstrated in experiments involving the transplant of cells expressing an activated form of TGF-b type I receptor. These cells are able to adopt an endodermal identity even when grafted to the animal pole, which usually contributes to neuroectoderm; moreover, they are able to restore endoderm derivatives when transplanted into the endoderm-deficient mutant casanova (David NB et al., 2001). The importance of Nodal signalling is also underscored from experimental evidence provided from mouse studies. The knockout mice for the Nodal gene lack the primitive streak, which promote the mesoendoderm differentiation (Conlon FL et al., 1994). Another line of evidence that Nodal signal is involved in endoderm forma-

tion in the mouse comes from the analysis of the Cripto mutant, a cofactor in nodal signalling belong to EGF-CFC family member. This mutant among other abnormalities, lack definitive endoderm (Ding J et al., 1998).

The simplest pathway, which emerges from these independent studies, locates VegT or other related factors, at the top of transcriptional hierarchy. It activates and reinforces the Nodal signalling cascade but also induces the expression of several transcription factors such as Sox17, Mixer/Mix/Bix and the Gata family. Interestingly, in Xenopus two members of the Bix gene family, Bix1 and Bix4, and Sox17 have been demonstrated to be direct targets of VegT and endodermal inducer factors. The injection of mRNAs encoding Bix4 and Bix1 leads to rescue of the depleted VegT embryos (Tada M et al., 1998; Casey ES et al., 1999). Moreover, two T-box binding sites in the promoter region are necessary for the correct expression of Bix4 in endodermal precursor cells. Furthermore, high concentrations of Bix genes in an animal cap assay induce the expression of endodermal markers (Tada M et al., 1998; Casey ES et al., 1999). The expression of Sox17 starts before the mid-blastula transition and extends throughout the presumptive endoderm. Sox17 overexpression in animal caps assays induces two late endoderm marker as IAFABP and Xlhbox8 (Hudson C et al., 1997). The direct activation of Sox17 by VegT has been demonstrated through the injection of an inducible form of VegT (VegTGR) that is able to promote the expression of Sox17 even when the protein synthesis is inhibit by addition of cycloheximide (Clements D et al., 2003). Endodermally-expressed sox genes have also been isolated in mouse and zebrafish. The Sox17 mutant mouse shows profound defects in the gut providing further evidence of

the involvement of this gene in endoderm formation (Kanai-Azuma M et al., 2002). In zebrafish, Sox17 is specifically expressed in endodermal precursors at the onset of gastrulation. A novel member of SoxF subfamily encoded by the Casanova gene has been recently determined (Dickens T et al., 2001; Kikuchi Y et al., 2001). Casanova mutants (cas) completely lack Sox17 expression (Alexander J et al.1999). Another transcription factor, whose molecular function is conserved across the phyla, is Mixer (Hudson C et al., 1997; Henry GL et al., 1998; Dickens T et al., 2001; Kikuchi Y et al., 2001; Kanai Y et al, 1996). In Xenopus, Mixer is expressed exclusively in endodermal precursor cells at beginning of gastrulation and can activate endodermal genes in animal cap cells. Dominant interference studies, using Mixer-EnR chimeras, indicate that it plays a role in the regulation of endodermal gene expression, and that it acts upstream of Sox17 (Henry GL et al., 1998). Molecular and genetic studies in zebrafish have clearly established that bon (a mixer family member) acts upstream of Sox17. In fact, Bon expression precedes the onset of Sox 17 expression (Alexander J et al., 1999), in the bon mutant sox expression is absent, and 90% of endoderm cells are missing (Kikuchi Y et al., 2001). In the mouse a mix-type gene, Mnl, is expressed in the nascent primitive streak where the endoderm end mesoderm originate and thus might play a role in the formation of these germ layers (Pearce JJ et al.,1999); however there is no evidence that it is an orthologue of the mixer gene.

The Gata family of zinc finger transcriptional activators have been implicated in endoderm formation across the phyla. The vertebrate genome contains at least six evolutionary conserved Gata genes. Three of them (Gata1, 2 and 3) are involved in haematopoietic devel-

opment, while Gata4, Gata5 and Gata6 are expressed in heart, extraembryonic tissue and endoderm (Patient RK et al., 2002). In Xenopus, gata4 and gata5 are expressed early in response to VegT or to TGF-b (Yasuo H et al., 1999, Weber H et al., 2000, Xanthos et al, 2001) and the misexpression of XGata5 recruits ectomesodermal cells to an endodermal fate (Weber H et al, 2000). Further demonstration of the role of gata genes in endoderm formation in vertebrates has come from the analysis of the zebrafish faust (fau) mutation. Fau encodes zebrafish gata5 and is required for endoderm formation upstream of Sox17 (Reite JF et al., 2001). In addition fau/gata5 is sufficient to induce Sox17 expression and requires the function of the novel Sox protein cas to do so (Kikuki Y et al., 2001; Reiter JF et al., 2001). Genetic analyses in the mouse have so far failed to reveal a clear role for gata4, gata5 or gata6 in endoderm formation, despite their widespread and early expression in this tissue (Molkentin JD, 2000).

Thus, after an initial activation of Nodal signalling through VegT, the second step of endoderm differentiation involves the *Sox, Mixer* and *gata* genes. Mixer and gata are regulated by Nodal signalling, that in turn regulate the expression of *Sox17* (Henry GL et al., 1998). To date, the role of *Sox17* is not clear but it has been shown to interfere with the Wnt-pathway. The Wnt molecules and their effectors appear to play an important role in mesoendodermal differentiation. The activation of the frizzled receptor by WNT protein induces the down-regulation of the GSK-3β kinase and stabilises cytoplasmic β-catenin. Stabilized β-catenin accumulates in the cytoplasm and is transported in the nucleus; here, it exerts its function by binding to TCF/Lef (Barker N et al., 2000), a transcriptional factor containing

an HMG box. The complex, TCF/β-catenin, activates downstream genes, among these siamois, involved in mesoderm differentiation. Sox 17 is able to bind β-catenin and antagonise the response to Wnt signal; it has been speculated that the repression of this cascade may be important for endoderm differentiation (Zorn AM et al., 1999).

The involvement of the Wnt cascade in endoderm differentiation has also been demonstrated in C. elegans. MOM2 encodes a Wnt homologue, Mom2 is received in the EMS cells by the MOM5 protein, the C.elegans version of the frizzled receptor. The result of this signalling cascade is to down-regulate the expression of the *pop-1* gene (TCF homologue) in the EMS daughter destined to become the ES cells. In the pop-1 deficient embryo both EMS daughter cells become E cells. Additional components of this pathway includes WRM1, a b-catenin homolog and MOM4 a MAP3K, necessary for endoderm formation (Rocheleau CE et al., 1999; Meneghini MD et al., 1999). The transduction of Wnt/MAPK signalling results in a change of nuclear level of POP1; in the MS cell that does not receive the Wnt signal, the nuclear level of POP1 are high while the E cells which are closest to P2 cells (the source of Wnt signal) the nuclear level of POP1 is low (Lin R et al, 1998)).

Although the wnt molecules play a key role in switching endoderm/mesoderm fate of the EMS cell, the first regulator of early blastomere fate identified so far is SKN-1. SKN-1 is a composite bZIP/homeodomain transcription factor whose function is restricted exclusively to the EMS cell. In this cell, SKN-1 activates expression of *med-1* and *med-2*, both encoding GATA type factors. The direct effectors of med genes activity are the transcription factors end-1 and end-3 also encoding GATA-transcription factors. While SKN-1 and MED1/2

messengers are detected in EMS cell, indicating their role in early differentiation of EMS from P2 cells, the END1/2 messengers are only detected in E cell lineage that suggests their role in differentiation of E cell from MS. After the E cell is specified, two other additional GATA factors ELT-2 and ELT-7 become expressed. Loss of ELT-2 leads to degeneration of the gut, whereas the lack of ELT-7 has no effect (Maduro MF, et al.2002). However, GATA factors seem to play an important role in the later stages of endoderm differentiation.

The wnt signalling and the GATA transcription factor have been implicated in mesendoderm differentiation in sea urchin. The mesendoderm specification starts when β-catenin, responding to an unknown signal, accumulates in the nuclei of all mesendodermal precursors and forms a complex with TCF. This complex regulates the expression of the signalling molecule Wnt8, which in turn activates the β-catenin/TCF signalling creating a positive feedback loop. The β-catenin/TCF complex also activates several transcription factors including krox, otx and krl, a kruppel like factor as early response gene and later endodermal genes, as *GataE, foxA* and *fox*. At the early blastula stage GataE becomes the major endodermal regulator. Its inhibition by morpholino antisense oligo blocks the endoderm formation (Davidson E et al., 2002). At blastula stage the endoderm start to segregate from the mesoderm. As a response to β-catenin accumulation, a subset of cells in the centre of the presumptive mesendoderm start to synthesize the Notch ligand Delta; as a consequence the surrounding mesoendodermal progenitors, which express Notch on their surface, become mesodermal. The mesoendodermal progenitors that are not in contact with Delta-expressing cells become endodermal (Sherwood DR et al., 1999).

1.4 The endoderm differentiation in Ascidians

The Ascidians, one group within the tunicate are marine invertebrates that belong to the most primitive branch of chordate phylum, the Urochordates or tunicates. The swimming larva is comparable to the amphibian tadpole; they have axial notochord flanked by muscles cell in the tail, a dorsal neural tube and a ventral endodermal strand.

The endoderm of an ascidian's tadpole larva is a simple tissue comprising about 500 cells and their lineage is almost completely documented (Nishida H et al., 1983, 1985, 1987). As mentioned, endoderm differentiation, in ascidians, is an autonomous process. This autonomy is dependent on maternal factors or pre-determinants pre-localized in the endoplasm of eggs and early embryos. The distribution of the determinants in the cytoplasm was investigated by cytoplasmic transfer experiments involving the fusion of blastomeres and non-nucleated egg fragments (Nishida H, 1993, 1994; Kumano G et al., 1998). Recently, the importance of these still unknown factors has been confirmed. In fact, Kobayashi K and Nishida H (2001) demonstrated that blastomeres committed to muscle fate, can be induced to express the endoderm marker alkaline phosphates by cytoplasm transfer. This occurs even after the recipients are already restricted to one cell type, and after the expression of the tissue-specific gene is already initiated. To date, no maternal factor, with this feature, has been isolated.

The fate restriction of the endoderm precursor in Ascidians occurs at a very early stage. In an 8-cells stage embryo the blastomeres of the vegetal pole or the animal pole, are already specified to give rise to different cell-types (fig. 2).

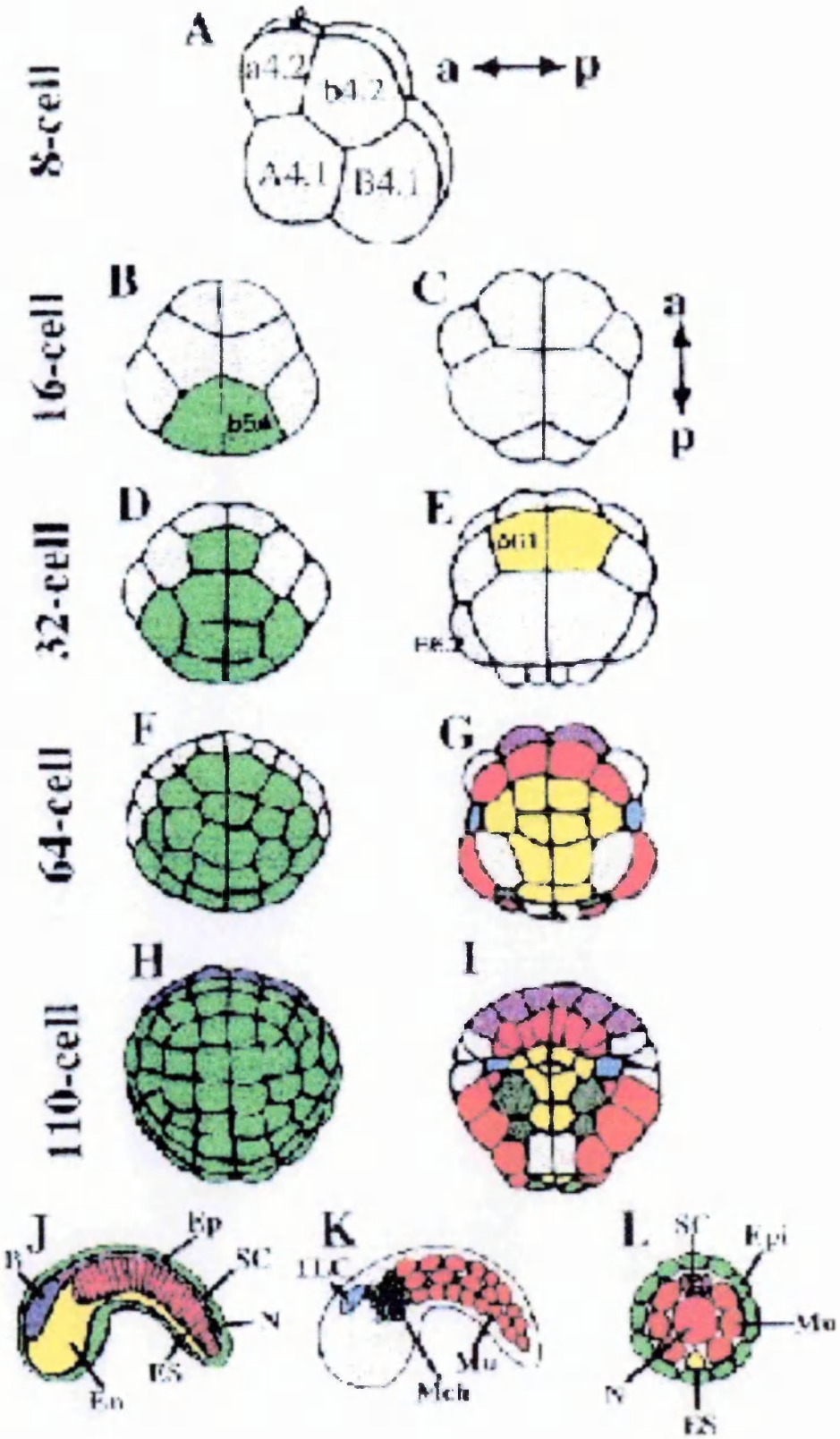


Fig.2: The Ascidian's cell lineage. This cartoon shows that the fate of each blastomere is determinated early. In particular, endodermal fate restriction occurs early in embryogenesis, at the 32-cell stage, when two pairs of vegetal blastomers become determined to give rise exclusively to endodermal cells. The process is completed before gastrulation, at the 64\110-cell stage embryo, when five blastomere pairs are restricted to the endodermal fate (Satoh N., Development Growth and differetiation,1996).

In an embryo at the 32-cells stage, almost all cells have undergone fate restriction. As far as endoderm is concerned, fate restriction occurs in at least two steps. Blastomeres A6.1 and B6.1, in a 32-cell stage embryo are already fate restricted, and will give rise exclusively to endodermal cells of a larva. Conversely, blastomers A6.3 and B6.3 will give rise to endoderm and also to other cell-types. After another round of division in a 64-cells stage embryo, five new blastomere pairs (A7.1, A7.2, A7.5, B7.1 and B7.2) appear that are restricted to endodermal fate.

To date, nothing is known about the factors that start the differentiation pathway of the endoderm in ascidians. Some data are emerging from the isolation of a maternal and a zygotic gene. The β-catenin is one of these maternal factors. Mis-expression of β-catenin in Ciona embryos induces the expression of endodermal markers, such as alkaline phosphatase, in cells fated to become notochord and epidermis. Conversely, the down-regulation of β-catenin by cadherin overexpression results in the loss of endoderm (Imai KS et al., 2000). Subtractive hybridization screens of mRNAs from embryos overexpressing either β-catenin or Cadherin has led to the identification of two endodermal genes a LIM-homeobox gene (*Cilhx*) and an NK2 gene (*Cittf1*). The expression of the Cilhx gene is detected at 32-cell stage in blastomeres A6.1, A6.3 and B6.1. At 64-cells stage its expression is inherited to their daughter cells: the endoderm, muscle and notochord precursor cells. This expression persists to the 110-cell stage and is then down-regulated. Later, neurulae, and tailbud embryos express Cilhx in a subset of notochord and brain cells. The β-catenin over-expression results in up regulation of Cilhx, while down regulation of Cilhx results in inhibition of

alkaline phosphatase expression; suggesting its role in early endoderm restriction pattern (Satou Y et al., 2001).

Cittf1 expression begins at 64-cell stage only in those blastomers committed to the endoderm fate. This expression persists until 110-cell stage, then disappears until the neurulae stage, to reappear in endodermal precursor cells at tailbud at larva stage; after metamorphosis Cittf1 is detected in the endostile. The mRNA injection of Cittf1 results in ectopic expression of alkaline phosphatase in blastomeres which normally give rise to notochord precursor cells (Ristoratore F et al., 1999), suggesting that Cittf1 is involved in the endoderm differentiation pathway. b-catenin over expression induces an up regulation of Cittf1 (Satou Y et al., 2001).

1.5 Aim

The aim of this project is to identify genes involved in the first steps of endoderm specification in *Ciona intestinalis*. We aim also at identifying genes that regulate expression of Cititf1, an early endoderm specific marker in Ciona.

Although the molecular mechanisms that rule endoderm differentiation are mainly conserved among different organisms they show some divergence. Therefore I did not limit my studies to the task of finding homologues of vertebrate genes involved in endoderm specification and decided to undertake two different and complementary strategies. In the first

one I tried to isolate genes already cloned in frog or in fish and that could also be involved in endoderm specification of *Ciona intestinalis*. In particular I attempted to clone the ascidian counterparts of Xenopus Sox17, a transcription factor containing a HMG domain and of Mixer, a homeobox gene.

In the second strategy, I focused my attention on the regulation of Cititf1, the earliest endoderm specific marker isolated thus far. Although its temporal and spatial expression seems to be finely regulated, the factors involved in its activation and likely function in endoderm specification are still unknown. Thus, in order to identify these factors I focused my studies on the regulatory region of Cititf1. We identified a sequence responsible for endoderm specific expression using mutational and biochemical approaches; this sequence was then used as bait in a one-hybrid screening in yeast to identify the factor(s) that to bind to it.

Chapter 2

Materials and Methods

2.1 Ascidian's eggs and embryos

Ciona intestinalis adults were maintained under constant light to induce oocyte maturation. Eggs and sperm were obtained surgically from the gonoduct. After fertilization, embryos were reared at room temperature in Millipore-filtered Seawater. Only the batches in which 90% or more of the eggs developed normally were used for experiments.

2.2 RNA isolation

Total RNA was isolated from eggs or embryos by the acid guanidinium thiocyanate-phenol-chloroform method (Chomczynski and Sacchi, 1987). The polyA+ RNA was prepared using Dynalbeads (OXOID) according to manufacturer's instruction. This polyA+ RNA was used to construct the DNA libraries.

2.3 cDNA library construction

The cDNA to construct the libraries was synthesized using the SuperScript choice System (Life Technology) according to manufacturer's instructions. In the case of Ascidians Sox 17 and Mixer counterpart screening, the cDNA was synthesized using oligodT primer from mRNA derived from 64-cell stage embryos and cloned into the Lambda ZAP vector system (Stratagene). After in vitro packaging approximately 10^6pfu were obtained. To perform the yeast one hybrid screening, the starting mRNA derived from 110-cell stage embryos, the cDNA was synthesized by random and oligodT primers, and cloned in pGAD424 plasmid in an EcoRI site (Clontech). This library was named as Gal4/cDNA-110. The fusion library was transformed in DH5α cells and has produced a total of $0.7*10^6$ cfu; the average size of about 800 bp, was estimate by PCR with Gal4-5' and Gal4-3' primers and the percentage of insertion was 70%. This library was amplified and collected in LB medium containing ampicillin at 50μg/ml. Subsequently the colonies were collected and the DNA was extracted using the QuiaQuick midi preparation Kit. The DNA concentration was about 2mg/ml. Before collecting, about 100 colonies were picked to re-analyse the percentage of insertion. This analysis revealed a decrease to 50%.

2.4 64-cDNA library screening

The hybridization of phage lifts was carried out at 58°C for 16h in a buffer containing

6xSSC, 5x Denhardt's solution, 0.1% SDS, and 100μg/ml denatured sonicated salmon sperm.

Full-length cDNA or partial cDNA encoding for the conserved domain (homeodomain for MIXER and HMG domain for SOX 17β) of Xenopus laevis MIXER and SOX 17β gene were used as probe. All probes were labelled by random priming using [α–^{32}P]-dATP and [α-^{32}P]-dCTP to a final specific activity of 2x10^{9} cpm/μg. Filters were washed at 50^{o}C in washing buffer containing 2xSSC, 0.1% SDS for two times 30 minutes and three times 60 minutes and exposed to X-AR Omat films (Kodak, Rochester, NY) with intensifying screens at -80^{o}C for 16h.

2.5 RT-PCR and PCR

The first strand was synthesized from total RNA of eggs and embryos at, 8-cell, 32-cell, 64-cell, 110-cell, gastrula, neurulae and larva stage using SuperScript first strand synthesis system for RT-PCR according to manufacturer's instructions (Life Technology). One-twentieth of the reverse-transcribed cDNA was used as template for polymerise chain reactions (PCR).

The PCR reaction was carried out using Taq polymerase (Roche) in standard reaction conditions with 2μM of each upstream and downstream degenerate primers, corresponding to the conserved region of HMG box of SOX family as described by Denny et al., 1992 and corresponding the first and third helices of paired like homeobox of MIXER family (n-mix5':CARMGIMGIAARMGIACNTT and nmix3': CTRTTYTGRAACCANACYTG). The reaction cycle

used was 94°C, 1 minute; 56°C, 2 minute; 72°C, 1minute. Successive annealing temperatures were scaled from 56°C to 44°C, with 3 cycles performed at each temperature. The PCR was also performed in standard conditions [(94°C, 1 minute; 46°C, 2 minute; 72°C, 1 minute) x 35 cycle followed by 10 minutes of extension at 72°C on 0.5 μg genomic DNA and on cDNA using primers corresponding to coding region for HMG domain of genomic fragment.

2.6 Southern blot

After amplification the PCR products were run in a 1% agarose gel containing 0.5X TBE and blotted onto nylon filter (Hybond N+; Amersham) according to manufacturer's instruction. The filter was hybridized, as described for library screening, using the HMG domain of Xenopus Sox17β or using oligonucleotide SOX-int in the same buffer at 45°C for 16h. The filters were washed at 45°C in washing buffer containing 2xSSC, 0.1% SDS three times for 30 minutes and exposed to X-AR Omat films (Kodak, Rochester, NY) with intensifying screens at -80°C for 16h

2.7 Colony hybridization

The RT-PCR product was cloned in TA topocloning vector according to manufacturer's instruction and transformed in TOP10 elettrocompetent cells. About 10^5 clones were

screened. The positive clones were isolated by colony hybridization. The hybridization of bacterial lift was carried out in the same condition described for the Southernblot analysis using the SOX-int oligonucleotide (GTATCCACACTTACACAACGCCGAACTC, fig. 4).

2.8 Preparation of the constructs :M15-20, M19.5, r1, r2, 5xS19pBS, 5xS19pLacZ, 5xS19pHisi1, 5xS19pBSLacZ and 5xS19E1bpBSLacZ

Mutant vectors M15-20, M19.5, r1 and r2 were made from plasmid #5E by site directed mutagenesis using the appropriate primers and following the Quick-Change Site Directed Mutagenesis Kit (Stratagene) instructions. Each primer contains the desired mutated base residues as instructed by the manufacturer. The mutation introduced consisted of transversion (A/C, T/G). Sequencing both strands using the dideoxy-termination method checked the plasmids.

The 5E plasmid consists of the genomic sequence of the Cititf1 promoter, which extends 365bp upstream to transcription start site, fused to lacZ reporter gene; this transgene is able to drive the expression of a lacZ reporter essentially in the same cells as the endogenous gene and it has been used as wild type in all electroporation experiments.

The concatenameric S19 oligonucleotide was synthesized by an Applied Biosynthesis instrument. After annealing, in 600 mM NaCl, the double stranded oligonucleotide was cloned in EcoRI/ BamHI pBluScript plasmid. This S19pBS plasmid was used to prepare the

constructs used for the electroporation and the yeast one hybrid experiments. Construct S19pLacz was obtained by digestion of construct S19pBS with XbaI, followed by Klenow fill-in and HindIII digestion. The fragment was run on agarose gel and purified by QiuaQuick purification kit according to manufacture's instruction. The ligase reaction was conducted at 16°C over night in presence of plasmid pLacZ (CLONTECH) pre-digested with Hind III- SmaI.

Constructs S19pHis was obtained by digestion of the S19pBS construct with EcoRV-XbaI. The fragment was run on an agarose gel and purified as above. The ligase reaction with plasmid pHis 1 (CLONTECH) pre-digested with Sma/XbaI was conducted at 16°C over night.

The construct S19pBSLacZ (for electroporation) was obtained by digestion of construct S19pBS with XbaI, followed by Klenow fill-in and XhoI digestion. The fragment was run on an agarose gel and purified as above. The vector pBS LacZ was obtained by digesting ESE-pBSLacZ with SphI followed by klenow fill-in and Xho digestion. The vector was purified from the ESE element by electrophoresis and gel purification as above. The ligase reaction was conducted at 16°C over night with a ratio 1:10 plasmid / insert.

The S19E1BLacZ construct was obtained by digestion of construct S19pBS with XbaI/XhoI digestion. The fragment was run on an agarose gel and purified as above and inserted in digested XbaI/XhoI E1BCAT. Subsequently the CAT reporter gene was eliminated by SmaI/ HpaI digestion and substituted with the blunt fragment of Cittf1 minimal promoter cloned upstream of LacZ reporter gene.

2.9 Electroporation

Electroporation of fertilized and Pronase E (Sigma) dechorionated *Ciona intestinalis* eggs was perfomed as described by Corbo et al. (Corbo J et al., 1997), except that the final volume in the cuvette was 700 µl and the capacitance setting 800 µF so that pulse range was 14-20 milliseconds. Embryos were allowed to develop at 18-20°C on 0.9% agarose-coated dishes in sea water until the desired developmental stage, then fixed in 1% glutaraldehyde in sea water for 20 minutes at room temperature, washed twice with PBS 1X and stained at 30-37°C in Staining solution (3 mM K3Fe (CN) 6, 3 mM K4Fe (CN) 6, 1mM MgCl2, 0.1% Tween20 and 200µg/ml X-gal in PBS). The analysis of the different constructs was carried out on the same batch of electroporated embryos, which were stained in parallel and for the same time length. For each construct a minimum of 30 embryos were analyzed. When not otherwise specified more than 90% of the embryos showed the same colour intensity.

2.10 Nuclear extracts

Crude nuclear protein extracts were prepared from *Ciona intestinalis* embryo (32-cell stage and 110-cell stage) and larvae. All steps were performed on ice and with ice-cold reagents. In brief, embryos were collected by centrifugation at 2000 rpm for 5 minutes in a Beckman JA (rotor JA-20) and washed two times. The pellet was collected by centrifugation and resuspend-ed in 5 times the pellet volume of solution I (10 mM Hepes, pH 7.9, 20 mM NaCl, 20 mM

MgCl2, and 1 mM DTT, 2.5 mM EGTA, 0.5 mM PMSF, 0.35 M sucrose), and left 10' on ice. After centrifugation the pellet was resuspended in 2 time pellet volume of solution I and homogenised using a glass Dounce homogenizer (type B pestle, 20 up-and down strokes by). The sucrose restore (9ml 75% sucrose + 1ml 10x solution) solution was added (1/10 of the total volume) and after mixing in the Dounce with pestle B, the nuclear fraction was then collected by centrifugation in pre-chilled Beckman rotor JA-20 at 10000rpm for 30". The pellet was resuspended in 1 time pellet volume of solution I plus sucrose. Spined as above. The nuclear pellet was resuspended in 5x the nuclear pellet volume of Solution II (10 mM Hepes, pH 7.9, 400 mM NaCl, 1.5 mM MgCl2, 0.1mM EGTA, 0.5 mM DTT, 0.5 mM PMSF, 5% glycerol) and subjected to mixing for 30 minutes at 4°C. Following centrifugation at 35000 rpm for 30 minutes in a Beckman L8-70M ultracentrifuge (in a VT-65Ti rotor) the concentration of extracted nuclear protein in the recovered supernatant typically was 1-4 mg/ml. The extracts were stored at -80°C in aliquots after dialysis against 500 volumes of nuclear dialysis buffer (0.2M Hepes pH7.9, 1M Kcl 2mM EGTA, 2mM EDTA and 20% of glycerol).

2.11 Gel shift assays

2 picomole of single strand oligonucleotide was labeled using T4 polynucleotide kinase (New England Biolabs) and [α-32P] ATP (Amersham), annealed and purified on Sephadex G-25 column (Boehringer Mannheim)

The binding reaction was carried out in a buffer containing 10 mM Tris pH 7,5 @ r.t., 60 mM KCl, 1 mM DTT, 1 mM EDTA, 13% glycerol and 2μg poly(dIdC). The mixture, including 4 μg of nuclear proteins, was incubated for 10 minutes before addition of (8x104 cpm) labeled probe in a final volume of 15 μl. After incubation for 20 minutes at room temperature samples were separated on 5% polyacrylamide/1X TGE gel and analyzed by autoradiography.

2.12 Yeast competent cell preparation and plasmid transformation

The yeast competent cells YM4271 (MATa, ura3-52, his3-200, ade2-101, ade5, lys2-801, leu2-3, 112, trp1-901, tyr1-501, gal4D, gal8D, ade5::hisG) and the transformation procedure were made performed following the Yeast Protocols Handbook (Clontech) with some minor modifications. Several colonies were inoculated in 50 ml of YPD (20 g/L Difco peptone, 10 g/L Yeast extract pH to 6.5, adenine hemisulfate with final concentration is 0.003%, and glucose to 2%) or SD (6.7 g/L Yeast nitrogen base without amino acids, pH to 5.8, 20 g agar for plates only, 100 ml of the appropriate sterile 10X Dropout Solution and glucose to 2%) and incubated at 30°C for 16–18 hr with shaking at 250 rpm to stationary phase (OD600>1.5). 30 ml of the overnight culture were transferred to a flask containing 300 ml of YPD at concentration of 0.2–0.3 O.D and incubated at 30°C for 3 hr with shaking (230 rpm). When the culture reached 0.4–0.6.OD, it was stopped, centrifugated at 2500 rpm for

5 min at room temperature. The cell pellets were resuspended in sterile distilled H2O and pooled into one tube in a final volume of 25–50 ml. The cells were collected by centrifugation as described above and the pellet was resuspended in 1.5 ml of freshly prepared, sterile 1X TE/1X LiAc. The efficiency of these competent cells was about 10^4-10^5 but decreased to 10^2 when a linearised plasmid was transformed.

In order to integrate the reporter vector, 10μg of linearized plasmid DNA and carrier DNA (in equal molar ratio) were transformed in 100μl of freshly prepared competent cell.

After mixing, 0.6 ml of sterile PEG/LiAc solution was added to each tube, vortexed at high speed for 10 sec to mix and incubated at 30°C for 30 min with shaking at 200 rpm. 70 μl of DMSO were added to the samples. The cells underwent a heat shock for 15 min in a 42°C water bath. Following, the cells were chilled on ice for 1–2 min and collected by centrifugation for 5 sec at 14,000 rpm at room temperature. The pellet was resuspended in 0.1 ml of sterile 1X TE buffer and plated on selective medium (SD/-His or SD/-ura).

For large scale screening, we used 0.1 ml of competent cells for each microgram of DNA from the library was used.

2.13 Genomic DNA isolation from yeast cells

The genomic DNAisolation from yeast cells was performed following the protocol from

the Molecular Biology handbook (Cold spring Harbour). Briefly, one colony was inoculated in 10ml of appropriate selective medium for 18 hours (until stationary fase O.D.>1.5) at 30°C, at 200 rpm. The cells were collected by centrifugation at 6000 rpm for 10 min at room temperature in a tabletop centifuge. The cell pellets were resuspended in 200ml of lysis buffer (2% (v/v) triton X-100, 1%(v/v) sodium dodecyl sulfate, 100mM NaCl, 10 mM Tris-HCl pH8, 1mM EDTA pH 8). Following, 200 ml of glass beads were added and the suspension was vortexed for 3' at higest speady. After addition of 200ml of TE the suspension was centrifuge to high speedy and precipitated with 1ml of ethanol.

2.14 Plasmid isolation from yeast cell

The plasmid was extracted from yeast cells according to clontech protocol.'s. Briefly, one colony was inoculated in 1 ml of appropriate selective medium for 18 hours (until stationary fase O.D.>1.5) at 30°C, at 200 rpm. The cells were collected by centrifugation at 6000 rpm for 10 min at room temperature in a tabletop centifuge. The cell pellets were resuspended in 50ml of potassium phosphate buffer (67mM KH2PO4;pH7.5). Following, 10 ml of lyticase solution was added and incubated for 30' to 37°C. After addition of 10 ml 20% the sample was vortex vigorously and the plasmid was purified using CROMA-SPIN 1000.

2.15 Preparation of digoxygenin DIG-labled antisense transcripts and In situ hybridization

The templates to synthesize RNA probes were prepared from the cDNApGAD clones, which were positive in the b-gal assay. These clones were amplified by PCR, with the Gal45' and Gal43' primers and subcloned in TA topo cloning vector (Invitrogen). Labelling of anti-sense RNA probes with digoxygenin was prepared using T7 and Sp6 RNA polymerases, according to the protocol supplied by Roche.

Tailbud and 110-cell stage embryo were fixed with 4% Para formaldehyde in 0.5M NaCl, 0.1 Mops (3-N morpholino propansulfonic acid), pH 7.5, at room temperature for 90 min-utes, dehydrated in 30%, 50% and 70% ethanol series and stored in 70% ethanol at -20°C. To perform the hybridization the embryos were washed with phosphates buffered saline (PBS) and manually deprived of envelopes. The specimens were treated with 2mg/ml of pro-teinase K in PBT (PBS containing 0.1% of Tween 20) (45 minutes, 37°C) and then post-fixed with 4% paraformaldehyde in PBS for 1 hour. After 1hour in pre-hybridization buffer (50% formamide, 5x SSC, 50µg/ml heparin, 50 µg/ml yeast tRNA, 5x Denhardt's solution, 0.1% Tween 20) at 55°C embryos were hybridized with 1 ng/µl digoxygenin DIG-labled antisense transcripts. Hybridization was carried out at 55°C for 18 hours. The specimens were washed in 50% formamide, 4xSSC, 0.1% tween 20 (2x15 minutes, 55°C) then in 50% formamide, 2xSSC, 0.1% tween 20 (2x15 minutes, 55°C) and in solution A(0.5 M NaCl, 10mM Tris-HCl , pH 8, 5 mM EDTA, 0.1% Tween 20) (3x20 minutes, 37°C), then treated with 20µg/ml RNase in solution A (30 minutes, 37°C) and washed with 50% formamide,

2xSSC, 0.1% tween 20 (20 minutes, 55°C), 50% formamide, 1xSSC, 0.1% tween 20 (2x15 minutes, 55°C), 1xSSC:1xPBT(15 minutes),PBT (4x5 minutes). RNA hybrids were detected immunohistochemically. After blocking in 5% normal sheep serum (NSS) in PBT (30 minutes), the specimens were incubated with 1:2000 alkaline phosphates- conjugated anti-DIG antibody (Boehringer Mannheim) in PBT (overnight, 4°C). The specimens were washed with PBT four times for 30 minutes and alkaline phosphates buffer (100mM NaCl, 50mM MgCl2, 100mM Tris-HCl , pH 9) (2x10 minutes). Signal detection was performed in APB containing 4.5µl/ml NBT (nitroblue tetrazolium salt) and 3.5µl/ml BCIP (5-bromo-4chloro-3-indolyl phosphate) according to supplier instructions (Boehringer Mannheim DIG RNA detection kit). When satisfactory signal over the background were obtained, the solution was replaced with PBT.

Chapter 3

Results and Discussion (i)

CiSoxE, the ascidian counterpart of Sox genes belong to the E subgroup

3.1 Summary

The HMG transcription factors and the Mix-like homeodomain proteins have been implicated in endoderm specification in Xenopus. In order to isolate the *Ciona intestinalis* counterpart of these molecules, I performed a PCR on Ciona intestinalis genomic DNA using degenerate primers to the Xenoups Sox17 and Mixer genes . This strategy led to the identification of an HMG containing protein, named CiSoxE. This approach did not allow identifying any Mixer gene homologues. CiSoxE shows high homology with members of SoxE group, which includes Sox8, Sox9 and Sox10. Ci-SoxE is expressed in the nervous system and in mesenchyme precursor cells; no transcript was detected in endoderm precursor cells.

3.2 Introduction

In recent years, a number of studies have begun to characterize the molecular determinants of the endoderm. Several studies, using different model systems, have defined a potentially conserved transcriptional pathway that serves to establish the vertebrate endoderm.

Nodal-related growth factors that belong to the TGFβ family, transcriptional regulators, such as Mixer homeodomain containing protein, and GATA factors have a very important role in endoderm commitment. These proteins cooperate to regulate expression of the HMG-box SRY-family transcription factors, the most conserved of which is Sox17 that may be an obligate mediator of endoderm development.

The fate restriction of endoderm cells in the Ascidians is an autonomous process; therefore it is less likely that secreted factors, such as the Nodal related factors, are involved in endoderm determination. In addition, although GATA factors are well conserved during evolution and seem to be functionally implicated in endoderm differentiation, the only member characterised so far in *Ciona intestinalis*, does not mark the endoderm precursor cells (D'Ambrosio P et al., 2003). Therefore, I decided to focus my efforts to isolate Mixer and Sox17 homologues in *Ciona intestinalis*.

3.3 Results

3.3.1 Cloning of Ciona Intestinalis SoxE

The first approach we undertook to clone the Ciona homologues of Mixer and Sox17 was to secreen a Ciona cDNA library. We contructed a cDNA library starting from polyA+ RNA derived from 64-cells stage embryos. Approximately $5x10^5$ phage plaques were screened with probes corresponding to Xenopus Sox17 HMG domain and Xenopus Mixer home-

odomain (see material and methods for probes and hybridization conditions). We could not identify any positive clones using this strategy.

Thus, we used an alternative approach to clone the *Ciona intestinalis* genes. I designed degenerated primers corresponding to conserved region of HMG box of the Sox family (as described by Denny P et al., 1992), and to the sequence encoding the first and third helices of the paired like homeobox of the MIXER family (see materials and methods). The RT-PCR and the PCR were respectively performed on total RNA at different embryonic stages and on genomic DNA.

The Sox primers led to the amplification from genomic DNA of a major band, approximately 500 base pairs in length (fig. 3a). This band hybridized with a cDNA probe encoding the HMG domain of Xenopus SOX17 (fig. 3b). It was purified, cloned and sequenced. The putative coding sequence, as shown in figure 4, displays a high degree of homology with HMG domain of some Sox genes. The Ciona sequence was named Ci SoxE. The genomic fragment of Ci SoxE contains two introns, 120 and 220 base pairs long (fig. 5); the presence of introns is also indicated by the consensus splicing acceptor and donor sequence highlighted in red . The region encoding the HMG domain is indicated

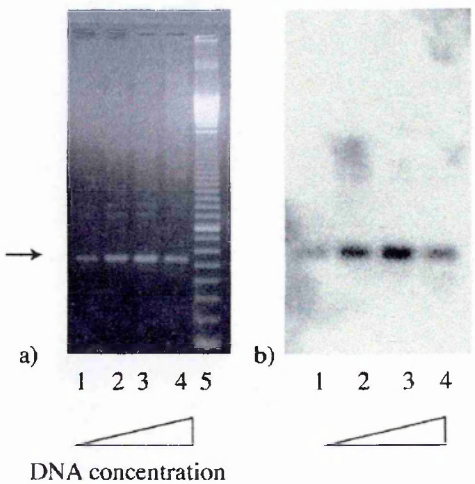


Fig. 3: ***Isolation of an HMG containing sequence from Ciona intestinalis genome.*** *a) PCR on genomic DNA, the amplified band corresponds to a fragment of 550 bp; b) the amplified samples were hybridised with the HMG domain of XSox17b. In each lane different amount of DNA were loaded (50-500ng). Lane 5 shows the MW markers (200 ng).*

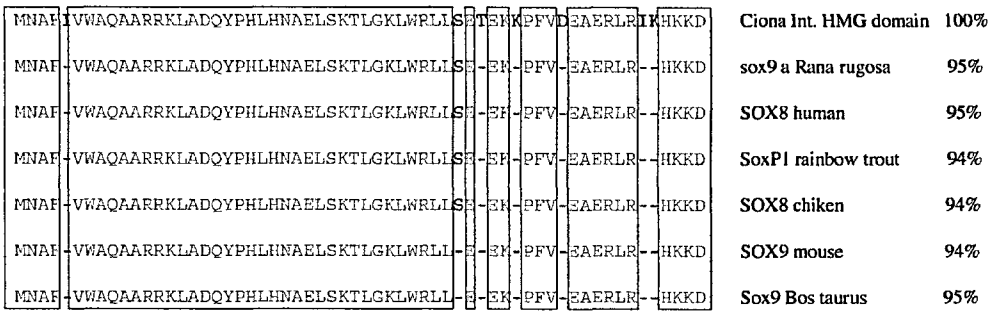


*Fig. 4: **Alignment of the HMG domain of Ciona Intestinalis with the HMG domain of the SOX gene belonging to the SoxE group.** The boxes indicate the conserved amminoacid while the "-" symbol indicates the amminoacids that are not conserved. The organism and the percentage of similarity are indicated on right of the picture.*

by the blue box in the figure. Three specific primers were designed on the genomic sequence of the Ci SoxE clone (see material and methods).

Two of these were used to perform an RT-PCR on total RNA derived from embryos at

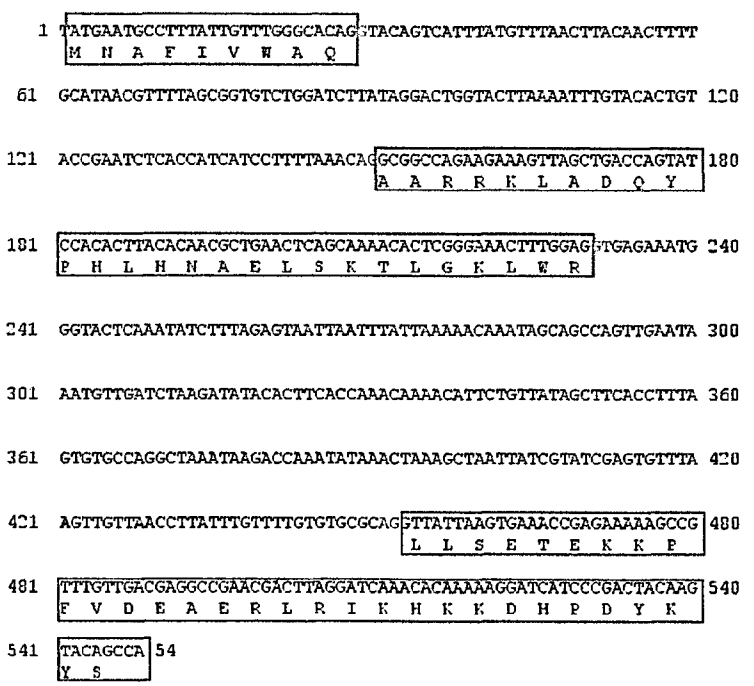


*Fig. 5: **The sequence encoding for the HMG box of Ciona Intestinalis SoxE.** The genomic sequence coding the HMG domain and the deduced amino acid sequence are boxed. The putative acceptor/donor nucleotides are indicated in red.*

different stages. The third primer, which was designed to the sequence encoding the HMG domain, revealed a 200bp band from the PCR products after hybridization (fig. 6 and material and methods for sequence). The PCR products were cloned in TA-topo cloning vector (INVITROGEN) and 10 clones for each PCR product were sequenced. All clones have shown 100% of homology with coding region of Ci SoxE. The cloning of homologue of MIXER, by same strategy, was unsuccessful.

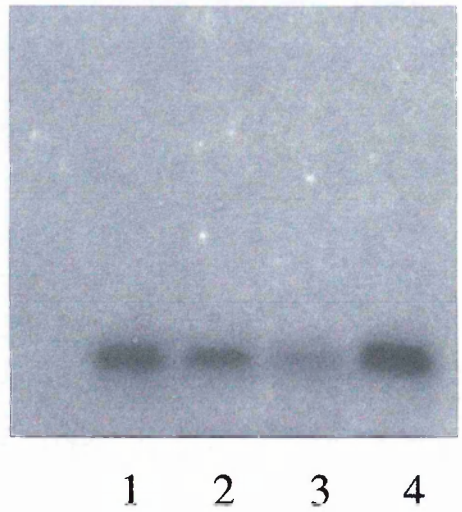


Fig. 6: ***Sox E is expressed at different embryo stages.*** *Southern blot of RT-PCR product on total RNA derived from 8-cell, 32-cell and larva stage embryos in lane 2, 3 and 4. In lane 1, PCR on pool of cDNA derived from different cell stages.*

3.3.2 Ci-SoxE transcript is detected in nervous system and mesenchyme precursor cells

The super family of Sox protein has been divided in ten subgroups named A to J (Wegner, M 1999; Bowles J et al., 2000). Members of the same subgroup usually share more then 80% amino acid identity within the HMG domain; the homology between different subgroups is less than 80%. In order to obtain a longer probe to be used in in situ hybridization experiments, the sequence of Ci SoxE was compared to the *Ciona intestinalis* cDNA database (ghost.zool.kyoto-u.ac.jp/indexr1.html) One clone showed 100% of homology with the deduced amminoacid sequence of HMG of CiSoxE. The full-length cDNA of this clone was obtained from Release I, a cDNA collection of *Ciona intestinalis* (provided from

prof. Nori Satoh), and was used for preparation of DIG-labled RNA probes (see material and methods).

In order to verify the site of expression of the Ci SoxE, whole mount in situ hybridization (WM-ISH) was carried out on a 110-cell stage embryo. A very strong staining was detected in the precursors of the nervous system cells, while a weaker signal was detected in the mesenchyme precursor cells (fig. 7). No expression was detected in endoderm precursor cells.

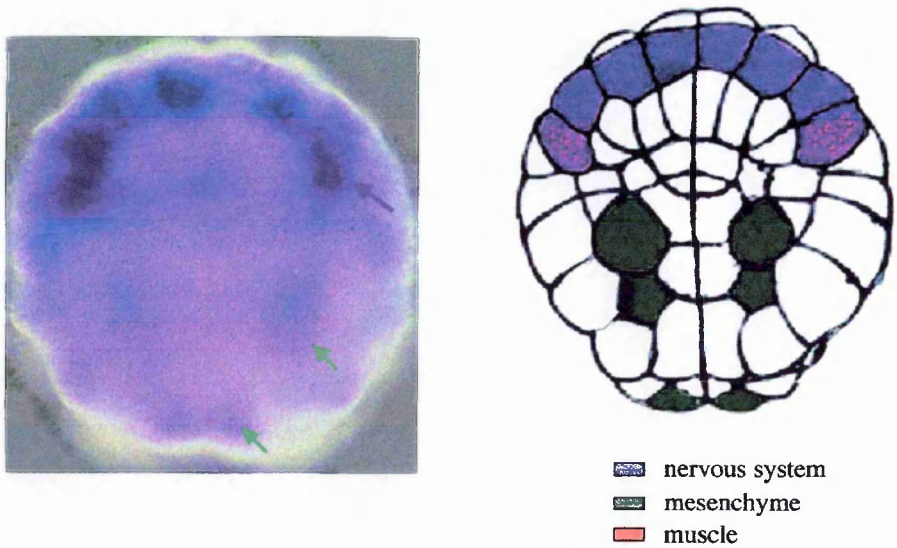


Fig. 7: ***In situ localization of Ci-Sox9 transcript in 110-cell stage embryo.*** *Whole-mount embryos were hybridized with digoxygenin labled antisense RNA probe and the signal was detected by histochemical staining with alkaline phosphatase. a) A representative 110-cell stage embryo; intense expression of the Ci-Sox9 transcript was detected in nervous system precursor cells while a weaker signal was observed in mesenchyme cells (The green arrows indicates the mesenchyme cells, the violet arrow indicates the nervous system precursor cells. b) cartoon of a 110-cell stage embryo representing the hybridization pattern seen in the picture.*

3.4 Discussion

Members of the SOX family transcription factors, characterized by the presence of a DNA-binding HMG domain, are found throughout the animal kingdom. The HMG domain

is present in large number of proteins that belong all to the SRY super family and only few aminoacid positions in this domain are conserved within the family. Thus, the HMG box can be highly diverse, but is strongly conserved in each subgroup (Wegner, M. 1999; Bowles, J. et al., 2000). Such a subgroup is a Sox protein family. The SoxF family is constituted by two isoform of Sox17, α and β, Sox7 and Sox18. The members of this family are found in mouse, in zebrafish and in Xenopous, and in most of case their expression were detected in endoderm precursor cells or in endoderm derivatives.

The Ciona HMG domain obtained by PCR both on genomic DNA and on cDNA, has a high percentage of homology with Sox genes that belong to subgroup E, but not to these that belong to SoxF subgroup. This family is constituted by SOX-8/9 and 10. Although for some SOX groups the conservation of sequence and structure correlates with similarity in function, members of the E group have very divergent functions. The Sox E members are, in fact, expressed in a range of tissues. Sox 10 is expressed in the neural crest, embryonic CNS and PNS, while Sox9 is expressed in the genital ridge, cartilage tissue, notochord, tubular heart structures (Takada S, et al., 2003; Moreno-Mendoza N, et al., 2003; Shea CM, et al., 2003; Mori-Akiyama Y, et al., 2003)

Thus, to establish if the CiSoxE gene has a role in early endoderm commitment or developmental processes, it is important to define its temporal and spatial expression. To this end, whole-mount in-situ hybridization on embryos at 110-cell stage was performed. The transcript of CiSoxE was identified in nervous and mesenchyme precursor cells, while no signal was detected in endodermal cells. These data show that the CiSoxE does not correspond

to an endodermally expressed Sox gene such as these belonging to Sox.

Moreover, only one Ciona HMG gene was cloned using degenerate oligonucleotides, a strategy that could have led to the identification of several HMG proteins. One possible explanation for this result could be that the small genome of Ciona, about 160Mb and 15,000 genes (Simmen MW et al, 1998), contains a less redundant set of genes compared to a vertebrate genome. In the case of Sox superfamily the molecular phylogenetic tree that Yamada and collegues have constructed (Yamada et al., 2003), have shown that only seven Sox subgroup have been identified in Ciona intestinalis and each subgroup is constitued by only one member in most of the case. In fact two Ci-SoxB1 and Ci-SoxB2 were assigned to SoxA/B/G group members; one Ci-SoxC to SoxC, one Ci-SoxD to SoxD, one Ci-SoxE to SoxE, one Ci-SoxF to SoxF and one Ci-SoxF to SoxF subgroup. Moreover, the Sox genes that belong to the group E and F are characterised by the presence of intron (Wegner M 1999; Bowles J et al., 2000), which could interfere with PCR amplification. These two points could explain the unsuccessful in the isolation of other Sox gene by using degenerates primers.

We can not exclude that the failure to identify a Ciona homologue for the Mixer gene is due to lack of sensitivity of the methods used. However, our results and the fact that the analysis of the *Ciona intestinalis* genome database does not identify a Mixer homologue either, strongly suggests that Ascidian's do not have a Mixer gene homologue.

Chapter 4

Result and discussion (ii)

Identification of an endoderm specific cis-regulatory element in the Cititf1 promoter

4.1 Summary

Cititf1 is an early and specific marker of endoderm development in *Ciona intestinalis* (Ristoratore F et al., 1999). A minimal promoter region of about 300 base pairs, which closely reproduces the spatial and temporal expression pattern of the endogenous gene, has been characterised (Fanelli A et al., 2003). Here, we have identified two repeats of eleven nucleotides, named R1 and R2, by mutational and in vivo assay and investigated their role in the regulation of the activity of the Cittf1 promoter. The R1 repeat seems to be involved in Cittf1 promoter regulation, although it is not essential. Interestingly, R2 contains an endoderm specific enhancer. Moreover, biochemical analysis showed that this region is specifically able to bind protein/s present in a nuclear extract derived from *Ciona intestinalis* embryo.

Furthermore, five copies of the R2 element are able to drive the expression of β-galactosidase reporter gene in endoderm cells, strongly suggesting an important role of this element in the endoderm specific expression of endogenous Cittf1.

4.2 Introduction

TITF1, otherwise known as TTF-1 or Nkx2.1 or T/EBP, is a member of the Nk2 family of homeodomain containg transcription factors. Several genes encoding an NK2-type homeodomain have been isolated from vertebrates. The role of TITF1 in development has been demonstrated in knock out mice; the titf1-null mice lack thyroid, pituitary and show severe defects in hypothalamus and lung, demonstrating an essential role of this transcription factor in the genesis of these organs (Minoo P et al., 1996).

Cititf1 is the ascidian's counterpart of mouse *Titf1* and represents the earliest endoderm specific marker isolated thus far. The mRNA encoding CiTTF1 is expressed during embryonic development in all endodermal precursors at the pregastrula/gastrula and larva stages (Ristoratore F et. al, 1999; fig. 8).

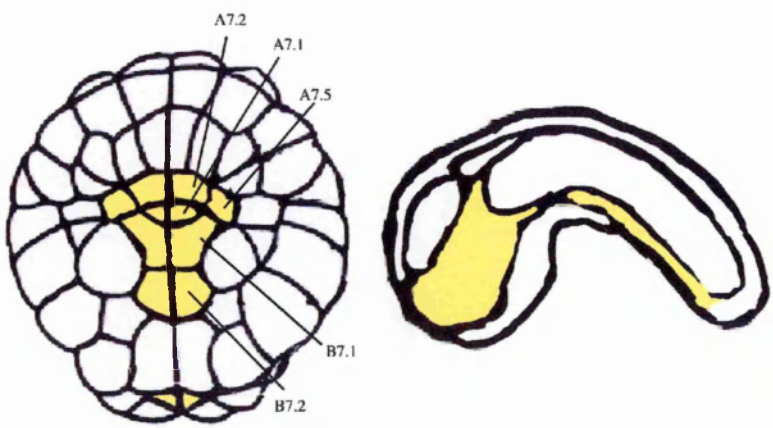


Fig. 8: ***Cittf1 is a marker of endoderm cells. Schematic representation of the hybridization pattern of the Cittf1 transcript at 110-cell stage and tailbud embryo;*** *the Cittf1 expression is first detected at pre-gastrula and gastrula stage in endodermal precursor cells; this pattern of expression is also maintained at early tailbud until the larva stage. The endodermal cells are stained in yellow.*

However, at the end of gastrulation the CiTITF1 mRNA becomes undetectable and reappears later at the middle tailbud and larva stage in anterior and ventral parts of head in endo-

derm tissue and endodermal strand. A genomic regulatory region extending 3,5 Kb upstream of the transcriptional start site showing a very tight endoderm specific control, has also been identified (Ristoratore F et al., 1999). Its characterization by deletion analysis led to the identification of a minimal region of Cititf1 promoter able to dictate endoderm specific expression (Fanelli A et al., 2003). Such endoderm specific regulatory region extends 367 bp upstream from *Cititf1* transcription start site and it promotes endoderm specific gene expression of a fused β-galactosidase reporter.

Further mutational analysis of the Cititf1 promoter led to the definition of a model that envisages the presence of positive and negative regulatory elements whose balance is responsible for *Cititf1* activation in the endodermal cell lineage and for its repression in the mesenchyme. In particular, this hypothesis comes from the expression of three constructs; the first one, containing the genomic fragment which extends -152bp upstream of transcription start site, shows a very strong β-galactosidase expression exclusively in the trunk mesenchyme, the second, which contains a genomic fragment extending -253 upstream of transcription start site, is capable to drive the reporter gene expression both in the endoderm and in the mesenchyme. The construct containing a genomic fragment, which extends -355 bp upstream of transcription start site drives the expression of the reporter gene only in endoderm cells. Furthermore, it has been shown that a fragment extending from -355 bp to -41 bp drives the expression of β-galactosidase in endoderm cells using the endogenous promoter region (fig. 9) or a heterologous promoter such as the Cihox3 basal promoter. These results suggest that these 300 base pairs contain enough information for endoderm expression.

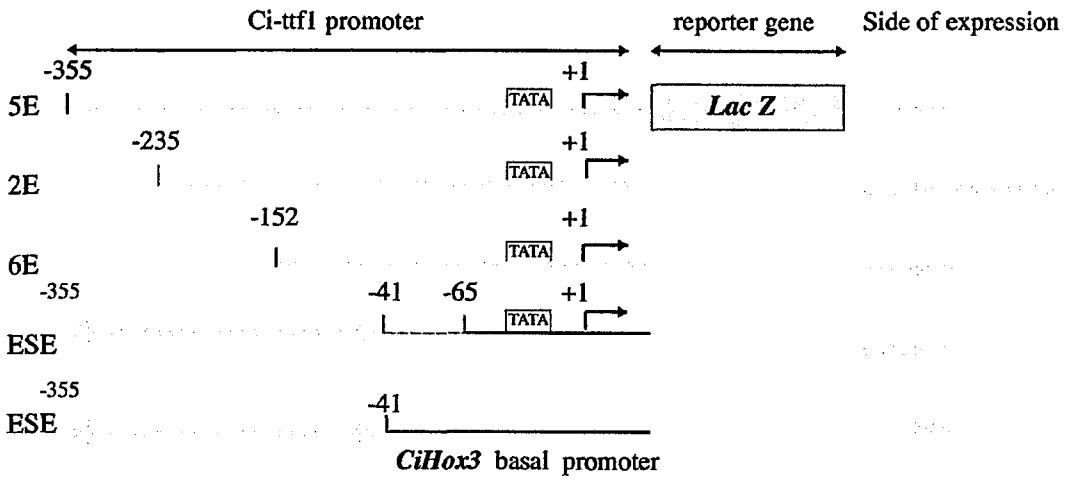


Fig. 9: ***Summary of the deletion analysis of Cititf1 endoderm-specific enhancer.*** *The scheme represents the Cititf1 promoter fragments analyzed in the electroporation experiments. The name of each construct is indicated on the left, the territories of reporter gene expression in the embryo are shown on the right (Fanelli A. et al., 2003). The ESE indicates endoderm specific enhancer element.*

Linker scanning mutagenesis has provided additional information on the mechanism of transcriptional activation in the endoderm. Out of 28 mutations introduced in the sequence most relevant for the endoderm expression, only three named M12, M19 and M20, resulted in a significant decrease of expression in the endoderm (fig. 10). However, no mutation resulted in the shutting off of this promoter, suggesting that its activity depends on the presence of redundant regulatory elements. Interestingly, the M19 and M20 mutations are very close and M20 falls in a putative binding site for a transcription factor containing an HMG domain (Harley et al., 1994; see below). These considerations prompted me to examine the region encompassing the two mutations, M19 and M20 more accurately.

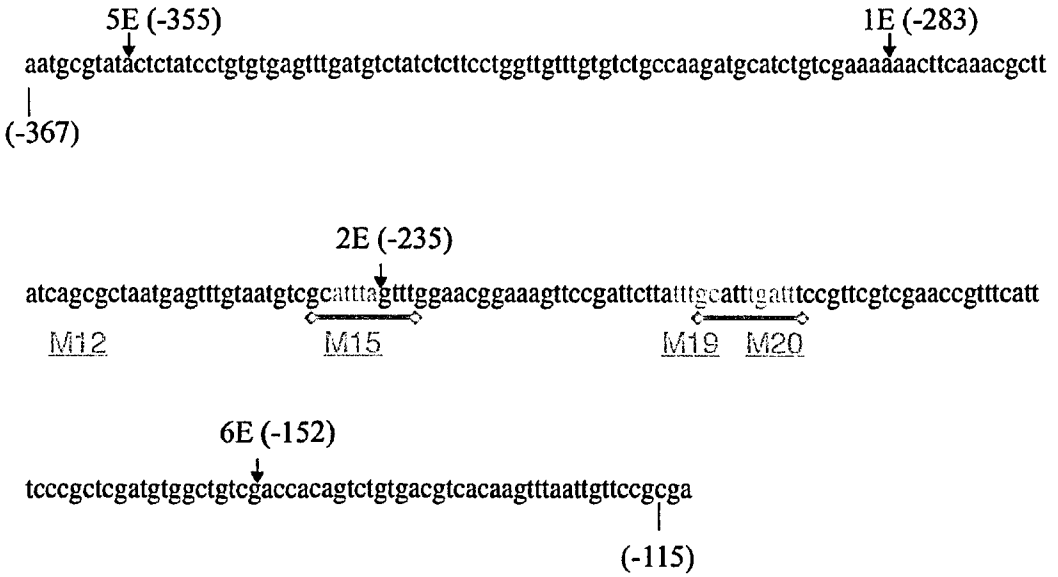


Fig. 10: ***Complete sequence of the Cititf1 enhancer region from -367 bp to -115 bp upstream of the transcription start site.*** *The three groups of five nucleotides named as M12, M19 and M20 and colored in orange, have beena shown to be able to reduce the expression of LacZ reporter gene in endoderm cells if they are mutated (Fanelli A. et al., 2003). Two repeated sequences (R1,R2) are underlined in red; within the R1 element five nucleotides are colored in green (M15) and their role is discussed in chapter 4.3.2. The deletion construct at the 5' end, whose expression patterns are shown in figure 9 are also indicate.*

4.3 Results

4.3.1 Identification and biochemical characterization of two repeats R1 and R2 in Cittf1 promoter.

The M20 mutation, which reduces the activity of the Cittf1 promoter (Fanelli A et. al.,2003) , falls within a sequence of eleven nucleotides (R2). A near identical sequence (with only 2 mis-matches) is repeated 26 base pairs upstream R2, and was defined as R1 (fig.10 underlined by red line). A search in a database for transcription factors revealed the consensus binding sites for HMG domain containing transcription factors (Kanai Y et al., 1996; Harley VR et al., 1994; fig. 11) within these repeats.

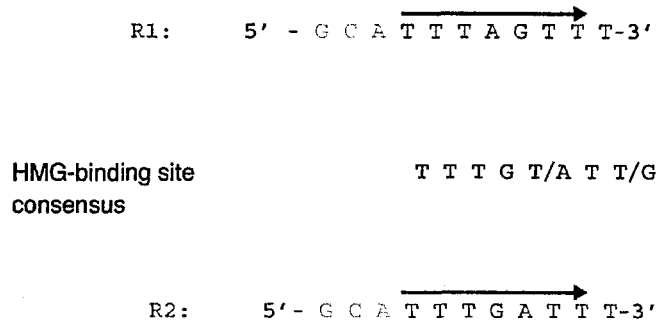


Fig. 11: ***The two repeats R1 and R2 contain a consensus sequence for HMG transcription factors.*** *The consensus binding site of transcription factors containing an HMG domain is indicated in the centre of the picture; the R1 and R2 sequences are respectively aligned with the HMG consensus binding site.*

The two sequences, R1 and R2, identify a repeat of eleven nucleotides, represented by the sequence 5'-GCATTTRRTTT-3'. In order to establish if both sequences contain elements able to bind to similar regulatory factor/s an EMSA was performed. Thus, two oligonucleotides, S1 and S2, containing respectively R1 and R2, and other four mutated oligonucleotides (S1m2, S1m3, S2m2 and S2m3) were designed for this assay (fig.12).

Fig. 12: ***Sequence of the oligonucleotides used for the biochemical characterization of the repeats contained in the ESE element.*** *The S1 and S2 oligonucleotides are 27 base pairs long and respectively contain the R1 and R2 repeats. Two mutated oligonucleotides for each repeat are also show; the m2 mutation covers the HMG consensus binding site while the m3 covers all nucleotides of each repeat. In (b) the sequence of S19 oligonucleotides is indicated and compared to S2.*

In detail, the S1m2 and S2m2 oligonucleotides are respectively identical to S1 and S2, except for the seven nucleotides corresponding to HMG-consensus binding site that were mutated The S1m3 and S2m3 oligonucleotides differ respectively from S1 and S2 because the whole repeated sequence has been mutated.

The double stranded oligonucleotides S1 and S2 were ^{32}-P-end-labelled and incubated with

protein extracts from *Ciona intestinalis* larvae. Five and four bands were detected using the S1 and S2 radiolabled double stranded oligonucleotides, respectively (fig.13 panel a). Four of them a, b, c and e, show the same mobility, suggesting that the two oligonucleotides could bind the same proteins, while only one, the d band, was detected with the S1 oligonucleotide.

The specificity of complex formation was tested using a molar excess of 100- fold of unlabeled, wild type and mutated oligonucleotides. The addition of unlabeled double stranded oligonucleotide S1 to the reaction interferes with the formation of the complexes, suggesting that they are specific (fig. 13 panel a lane 2). Interestingly, the addition of a molar excess of 100- fold of unlabeled S1m2 or S1m3 competitors, interfered with formation of the a, c and e complexes (fig. 13 panel a lane 3 and 4), indicating that the mutated nucleotides are not involved in formation of these complexes. The addition of the two competitors did not interfere with formation of complexes b and d, suggesting that the mutated bases are important for this binding.

The complexes formed upon incubation of ^{32}-P-end-labelled S2 double stranded oligonucleotide with crude extracts were completely displaced when the unlabeled double stranded oligonucleotide S2 was added (fig. 13 panel a lane 6). Also in this case, the addition of a molar excess of 100- fold of unlabeled double stranded S2m2 oligonucleotides interfered with formation of the complexes a, c and e while did not interfere the formation of complex b (fig. 13 panel a lane 8). Instead, a molar excess of 100- fold of unlabeled S2m3 did not interfere with the formation of any complex (fig. 13 panel a lane 9).

Taken together these results indicate that the eleven nucleotides of each repeat could be

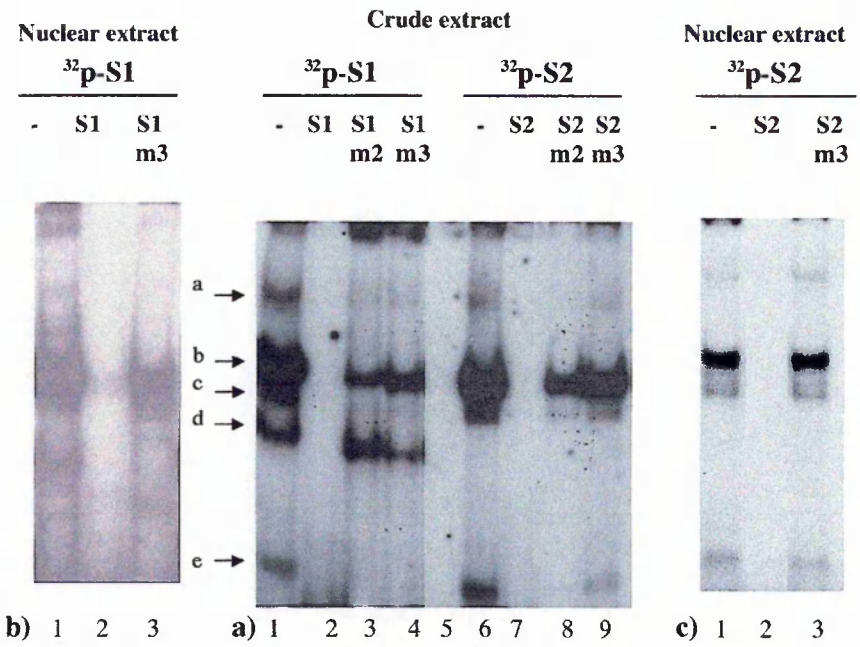


Fig. 13: ***The two repeats R1 and R2 contain DNA-binding sites involved in Ci-ttf1 promoter regulation.*** *The gel mobility shift assays were performed with S1, S2 oligonucleotides, crude and nuclear larva extracts. (Panel a) The radiolabled double-stranded oligonucleotide S1 and S2 (indicated as 32p-S1 and 32p-S2) are incubated with crude extracts derived from larva C. Intestinalis. Both S1 and S2 oligonucleotides are able to form several complexes indicated by the arrows and small letters (a, b, c, d and e). The cold probe S1, S1m2, S1m3, S2, S2m2 and S2m3 were used as competitors. The S1 and S2 probes were also incubated with nuclear extracts as shown, respectively, in panel b and in panel c. The cold probe S1, S1m3, S2 and S2m3, were used as competitors.*

involved in formation of the complex b, as suggested by using of the competitors S1m3 and S2m3, which contain the corresponding sequence mutated and are unable to displace the formation of the complexes. However, other nucleotides in S1 or S2 oligonucleotide are involved in binding of several proteins, as indicated by formation of the several complexes. Moreover, the lack of displacement of some them by using the competitors S1m2 or S2m2 that contain mutated bases in HMG- binding site consensus, suggests that an HMG-containing protein could be involved in their formation. Using nuclear extracts has also performed the biochemical analysis of R1 and R2 repeats and similar results have been obtained (fig.13 panel b and c).

In conclusion, several proteins that recognise these oligonucleotides are present in crude extracts from larvae of *Ciona intestinalis*. However, only some of these proteins bind to the

R1, R2 repeats. Therefore, before pursuing the biochemical analysis further, I decided to verify the functional role of the R1 and R2 sequences using an in vivo approach.

4.3.2 Mutation in R1/R2 site compromise the expression of theCittf1 promoter/lacZ transgene

Previous studies identified that a 367-bp DNA fragment from the Cititf1 5' regulatory region is sufficient to reproduce the Cittf1 expression profile. This region directs the expression of the β-galactosidase reporter gene in electroporated embryos to the endoderm of the trunk and to the endodermal strand (Fanelli A et al., 2003). The results obtained by linker scanning mutagenesis suggested the presence of redundant regulatory elements indicating that the cooperative interaction of different transcription factors could be responsible for Cititf1 promoter regulation. The identification of the two repeats R1 and R2 provides further support to this concept.

Two mutations M15 and M20 (fig. 10) fall into these repeats and are tested in a previous work (Fanelli A et al., 2003). Interestingly, while the introduction of the M20 mutation in wild type promoter (#5E) reduces the level of expression of β-galactosidase, no effect was observed after introduction of the M15 mutation (Fanelli A et al., 2003). The biochemical data, described above, suggests that each repeat is bound by the same factors. Then, I decided to test the effect of the combination of the two mutations M15-20 (fig 14).

The double mutant was prepared starting from the wild type plasmid #5E described in

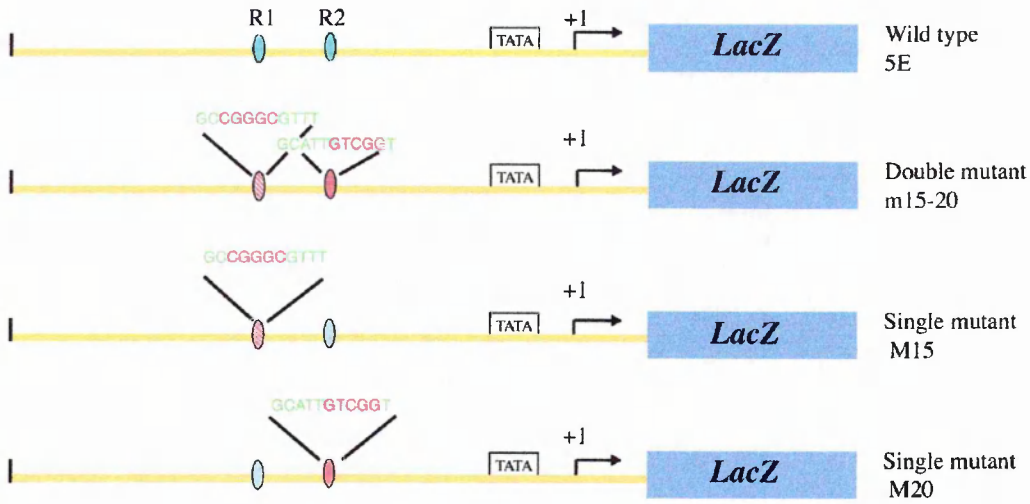


Fig. 14: ***Schematic representation of the mutations tested in vivo.*** *The wild-type R1 and R2 are indicated as green circles, the mutants in red. The respective mutated sequence in the R1 and R2 site, are indicate in red. The name of each transgene is indicated on the right of the construct. The m15-20: the 5E construct carrying the double mutation; the m15 and m20 are the plasmids carrying ofthe single mutations (Fanelli et al., 2003). All mutations were obtained by transversion. Both wild types and mutants are cloned upstream the Lac Z reporter gene and inserted in fertilized eggs by electroporation.*

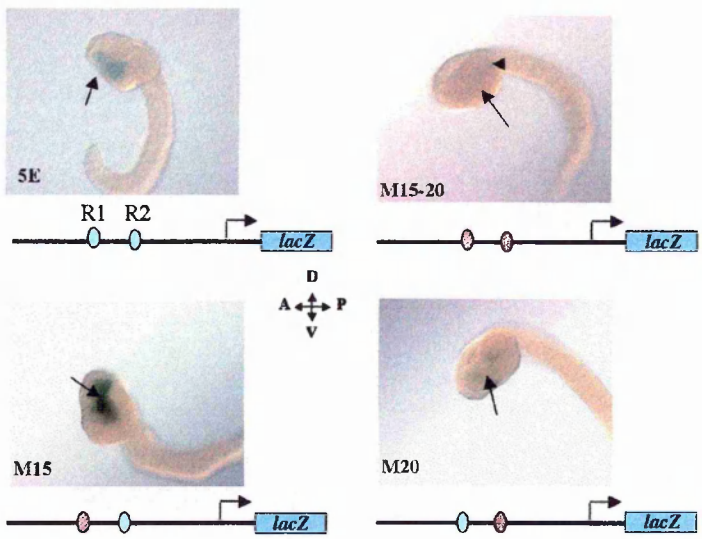


Fig.15: ***The double mutation M15-20 in R2 and R1 abolishes the endoderm activation of Cititf1 promoter.*** *A representative tailbud embryo electroporated with the plasmid containing the double mutation M15-M20 demonstrates the complete abolishment of reporter gene expression. The result obtained in the same experiment by introduction into Ciona embryos of the control construct WT (#5E) and the transgene carrying the single mutation M15 in R1 site and M20 in R2 site are also shown. The arrows indicate the endoderm tissue; arrowheads indicate the ectopic expression in mesenchyme cells.*

Fanelli A et al., 2003 and in material and methods. This and the following constructs were introduced into *Ciona intestinalis* fertilized eggs via electroporation. After different periods of

growth at 18°C, the embryos were collected, fixed and stained for β-galactosidase activity.

Figure 15 shows representative embryos electroporated with plasmid 5E, single mutant M15 or M20 and the double mutant M15-M20. The intensity of staining (observed in at least 30 embryos and in 3 independent experiments) shows that the double mutant is less active than the wild type (5E) and the single mutant (M20). Moreover, a very weak signal is detected in mesenchyme cells in 20 % of embryos electroporated with the plasmid carrying the double mutation. The expression in mesenchyme cells was also obtained in a low percentage of embryos when we electroporated the construct containing the genomic fragment that extends -152bp upstream to transcription start site. This data supports the model, previously hypothesized, that expression in endoderm cells results from the balance between endoderm activation and mesenchyme repression.

This result strongly suggests that the two repeats cooperate in endoderm specific regulation of the Cititf1 promoter.

4.3.3 The R2 site contains an endoderm specific cis-regulatory element

In order to understand if the diverse result obtained from analysis of single mutation M15 and M20 depends from the different bases changed in the two similar sequences, I decided to construct new mutations. All eleven nucleotides in each repeat were mutated in the context of the 5E wild type plasmid (fig. 16). Surprisingly, the mutation of all nucleotides in the R1 site has no effect in the expression of transgene. The embryos electroporated with this mutant (indi-

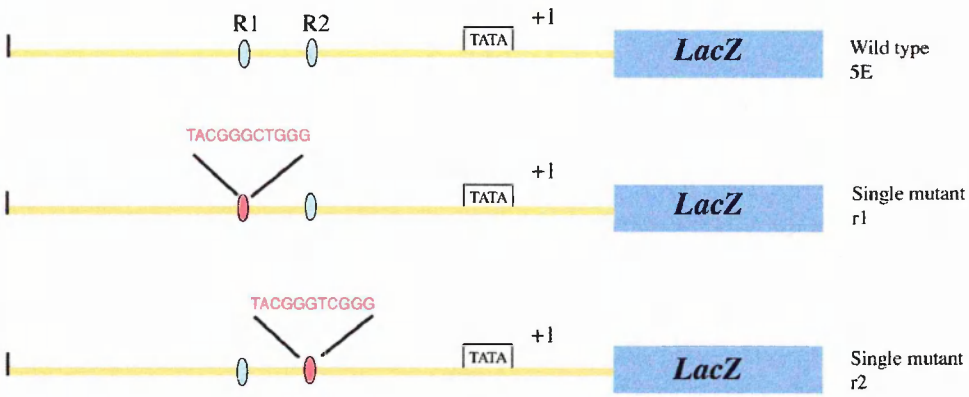


Fig.16: ***Schematic representation of the mutation tested in vivo.*** *The wild type R1 and R2 are indicate by the green circle; the respective mutated sequence covering the entire repeats and are indicate by the red circle; the mutations were obtained by transversion. In the r1 plasmid the whole R1 site was mutated, while in r2 the whole R2 site was mutated. Both wild type and mutants are cloned upstream of Lac Z reporter gene.*

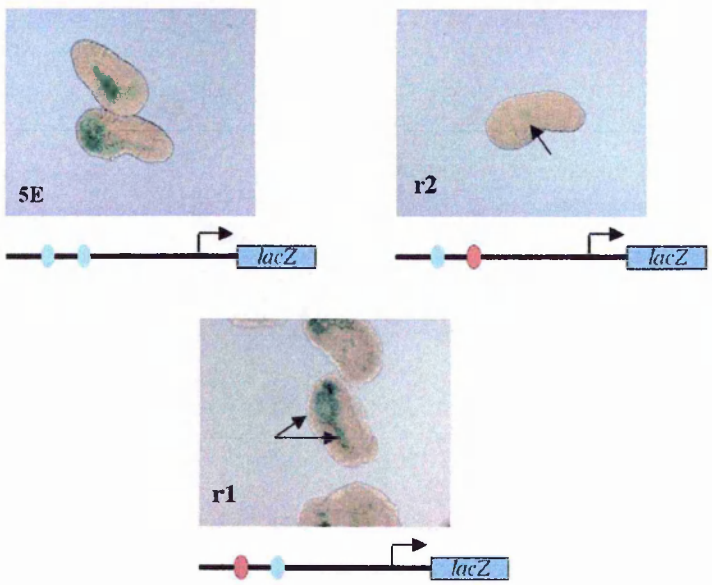


Fig.17: ***The mutation of the R2 site abolishes the endoderm specific expression of Cittf1 promoter.*** *The expression of the reporter gene LacZ in endoderm tissue was not detected in early tailbud embryo electroporated with the plasmid 5E mutated in the R2 site, but the ecopic expression in mesenchyme cells was detected (arrow). The result obtained, in the same experiment, after electroporation into Ciona embryos of the 5E wild type plasmid is shown as a positive control. The embryo electroporated with plasmid 5E mutated in R1 site (r1), shows the same phenotype of wild type 5E plasmid as indicates b arrows.*

cated as r1) show expression of the transgene in endoderm cell of the trunk and the endodermal strand, as seen with the wild type (#5E). In contrast the mutation that encompass the R2 site abolishes the expression of β-galactosidase in the structures expected (as shown in the figure 17). A weak signal was detected in mesenchyme cells in 20 % of embryos.

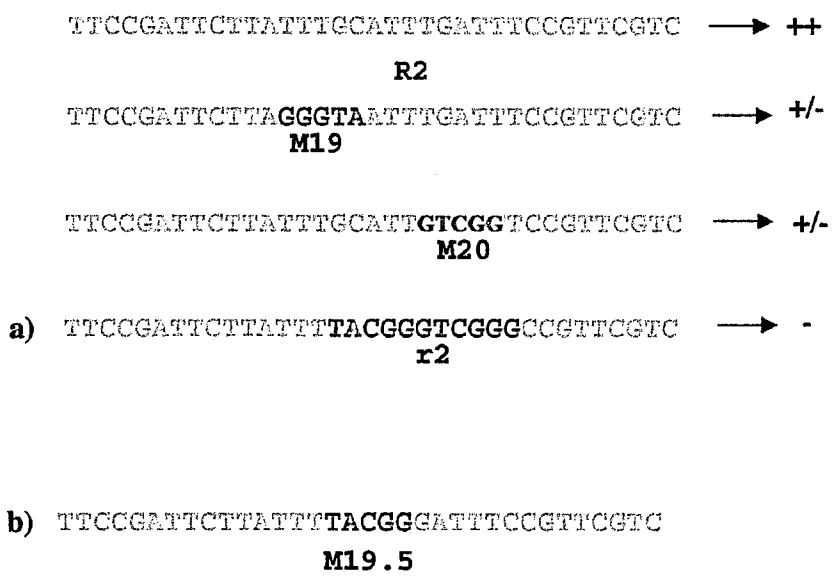


Fig.18: ***Summary of the mutation tested by electroporation in vivo.*** *a) The sequence of wild type R2 repeat and corresponding mutations M19, M20 and r2 are shown on the right of the picture; the site and level of expression, indicated as + or +/- are shown on the left of the picture. (b) indicates a new mutation located between M19 and M20 ,named M19.5*

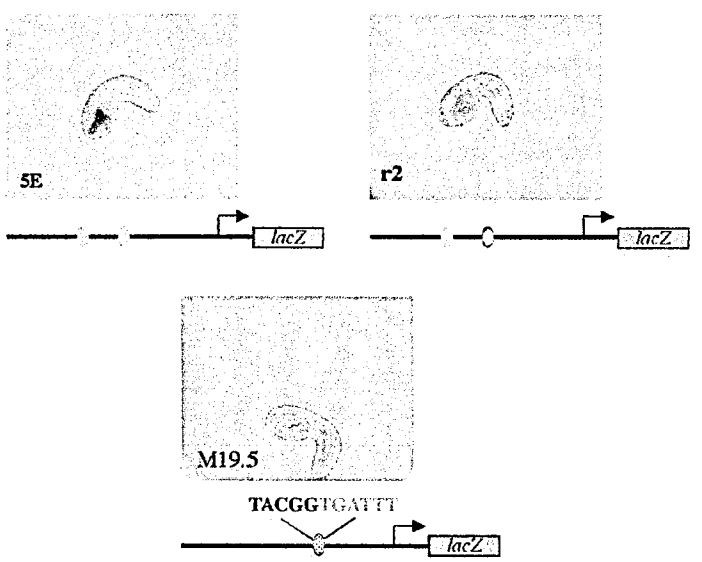


Fig. 19: ***Five base pairs in the 5' half of the R2 site are sufficient to drive endoderm activation of Cititf1 promoter.*** *Tailbud embryos electroporated with the plasmid carrying the M19.5 mutation completely lack expression of the reporter gene in endoderm cells. Also shown are the results obtained in the same experiment with the control construct 5E and the plasmid containing the mutation of whole R2 site.*

A summary of promoter activity shown by different mutations in the R2 region (fig. 18 a) suggest that important sequences are located between the M19 and M20 mutations. I decided to

construct a new mutation named M19.5 (fig. 18 b). This mutation abolishes promoter activity demonstrating an important role of this sequence (as shown in the figure 19). Furthermore, after a search in database for putative binding sites of known transcription factors I found that this sequence corresponds to a putative binding site for OCT3/4 transcription factor.

In conclusion, the R2 site has an essential role for promoter activity, which is probably supported from R1. Within the R2 site, five base pairs seem to play an important role in regulation of the promoter activity.

4.3.4 An oligonucleotide smaller than S2, is still able to bind nuclear protein/s

To study protein binding to the eleven nucleotides of the R2 site, a smaller oligonucleotide 19 bases long (S19) was designed. The sequence of the S19 oligonucleotide is shown in figure 20 a. In order to establish if this oligonucleotide was still able to form a complex, an EMSA was performed. The two double stranded oligonucleotides S19 and S2 were 32-P-end-labelled and incubated with larva nuclear extracts. Only one band was detected after incubation of the radiolabled double stranded oligonucleotides S19 with larva nuclear extract (fig. 20 panel b lane 1). The complex was displaced when an excess (100-fold) of cold S19 competitor was added. The addition of an excess of unlabeled double stranded S2m3 oligonucleotides or unrelated oligonucleotides (fig.20 b lane 2,3 and 4) did not interfere with complex formation. This strongly suggests that the formation of the complex depends on nucleotides within to R2 repeat. Conversely, multiple complexes were

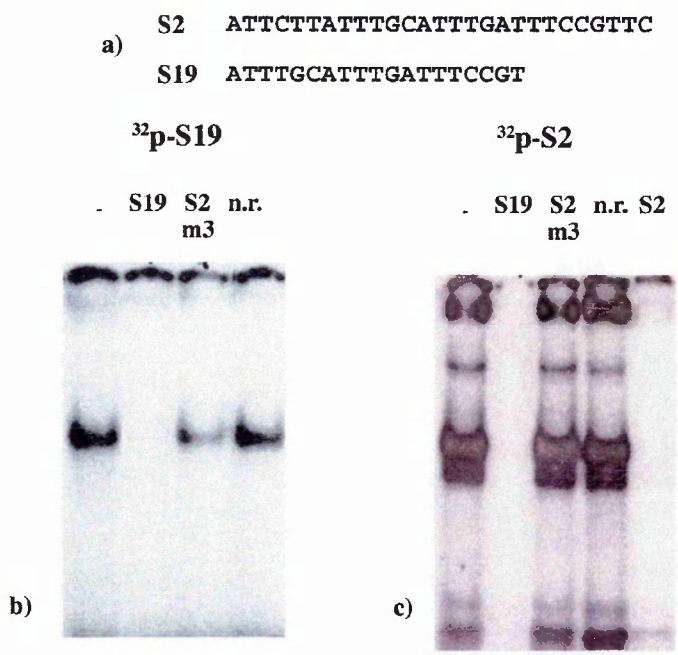


Fig.20 : ***One protein complex forms on the S19 oligonucleotide.*** *Panel a) Sequence of the S19 and S2 oligonucleotides. Panel b) The radiolabled double-stranded oligonucleotide S19, 19 base pairs long, was incubated with 110-cell stage nuclear extract, only one complex was formed. After addition of an excess (100-fold) of unlabeled double-stranded oligonucleotide S19, the complex was displaced. As a control, an excess of S2m3 or unrelated unlabeled double-stranded oligonucleotides were introduced into the assay and no effect was observed. Panel c) The six complexes formed upon incubation with Ciona intestinalis 110 cell stage nuclear extracts with the radiolabled double-stranded oligonucleotide S2 were displaced upon incubation with an excess of unlabled double-stranded oligonucleotide S19 as well as the unlabeled double-stranded oligonucleotide S2. As a control, an excess of S2m3 or unrelated unlabeled double-stranded oligonucleotides were introduced into the assay and no effect was observed.*

detected with the S2 oligonucleotides, and all were displaced when an excess (100-fold) of unlabeled S19 or S2double stranded oligonucleotides, were used as competitors (fig.20 panel c lane 2 and 5). The addition unlabeled double stranded S2m3 or unrelated oligonucleotides did not interfere with complex formation (fig.20 panel c lane 3 and 4). This data further confirms the result obtained with the S19 oligonucleotide.

Moreover, it suggests that the eleven nucleotides of the R2 site are responsible for the formation of the complexes obtained with a longer oligonucleotide. These results lead us to hypothesize a model where one protein binds to the repeat and in turn interacts with other proteins (see discussion). These results were further supported by data obtained in vivo.

4.3.5 The expression pattern of protein/s that bind to the R2 site correlates with Cititf1 expression

The EMSA analysis described was performed using nuclear extracts derived from embryo at the larva stage, when expression of Cititf1 is detected in all endodermal cells of the trunk. Because the expression of Cititf1 begins in endoderm precursors cells at 72-110 cell stage embryos, it is reasonable to suppose that the factors responsible for its regulation must be present during that developmental period. Therefore, an EMSA using nuclear extracts derived from embryos collected at different embryonic stages, was performed. The 32-P-end-labelled S19 double stranded oligonucleotide was incubated with same amount of nuclear extracts derived from embryos collected at 72-cell stage and 110-cell stage.

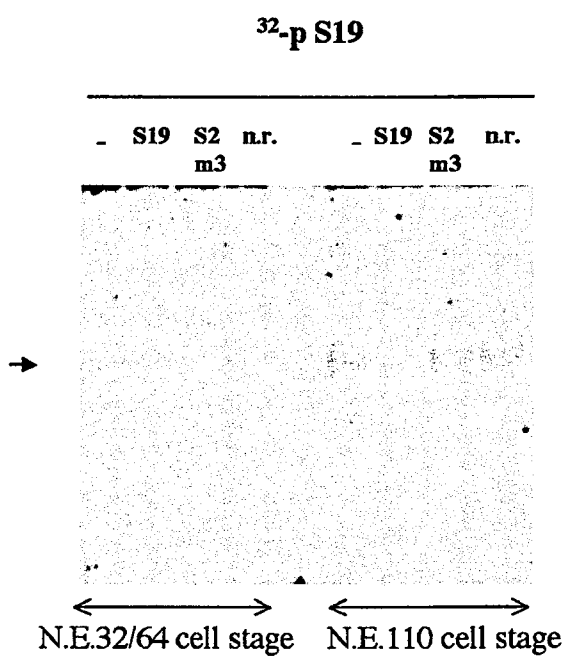


*Fig.21 : **The expression of proteins binding to R2 site correlates with the temporal expression of Cititf1** The radiolabled double-stranded oligonucleotide S19 was incubate with nuclear extract derived from 32/64 and 110 cell stages indicated by the double arrows. After incubation with 110-cell stage nuclear extract only one complex was formed. After addition of 100-fold molar excess of unlabled double-stranded oligonucleotide S19 the complex was displaced. The addition of 100-fold molar excess of unlabeled double-stranded oligonucleotide S2m3 or unrelated oligonucleotides had not effect. A very faint band is also detected at 32/64 cell stage.*

The EMSA analysis performed with nuclear extracts derived from 110-cell stage revealed the formation of a single complex (fig. 21 lane 6). Competition with increasing amounts (100-fold excess) of S2m3 or with unrelated controls did not alter the DNA-protein complexes (fig. 21 lane 8 and 9).

Furthermore, incubation of oligonucleotide S19 with nuclear proteins from the 32/64 cell-stage embryos resulted in the formation of a minor complex (faint band), suggesting that the protein responsible for endoderm specific activation of Cititf1 is present at this embryonic stage (fig. 21 lane1), although no further conclusion can be drawn.

4.3.6 Five copies of the S19 oligonucleotide are sufficient to drive the expression of the reporter gene LacZ in endoderm cells

In vivo and in vitro data obtained thus far allowed me to identify the R2 cis-regulatory region and to establish that it is necessary for activation of the Cititf1 promoter in endoderm cells. To establish if it is also sufficient to drive endoderm specific expression, it was essential to analyse its regulatory activity when inserted upstream a Cititf1 minimal promoter. The Cititf1 minimal promoter is the genomic fragment that extends only -65 bp upstream the transcription start site. When this fragment was used to drive expression of a reporter it resulted in unstained embryos (Fanelli A et al., 2003). Five copies head to tail of S19 oligonucleotide were cloned upstream of the Cititf1 minimal promoter. Moreover, to further investigate the role of the five bases identified to be important for the regulation of the activity of Cititf1 promoter (see in paragraph 4.3.3 about M19.5 plasmid), we developed a new oligonucleotide, S19m in which these five bases are mutated. Five copies of S19m were cloned head to tail upstream the Cititf1 minimal promoter (fig. 22).

The two constructs, named 5xS19pBSLacZ and 5xS19mpBSLacZ, were introduced into

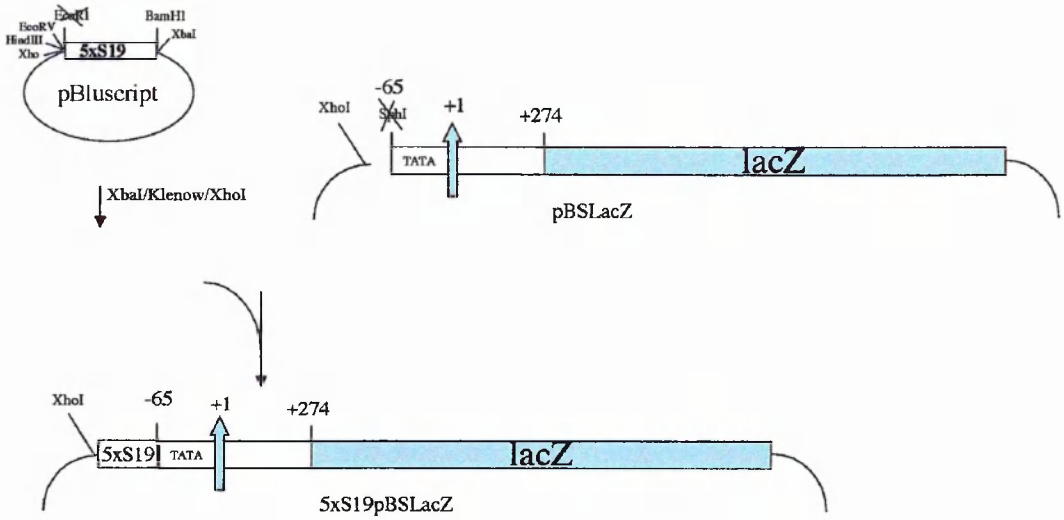


Fig. 22: ***Schematic representation of the strategy to prepare the construct containing the S19 concatamerised oligonucleotide.*** *The 5xS19 was cloned in pBS plasmid in EcoRI/BamHII and subsequently, in pBSLacZ 65 base pairs upstream the transcription start site (indicated as +1) of Cittf1 minimal promoter and its the TATA box. The pBSLacZ was obtained from digestion of ESEpBSLacZ; the ESE element was eliminated by gel purification. The same strategy was used to prepare the construct containing the m19.5.*

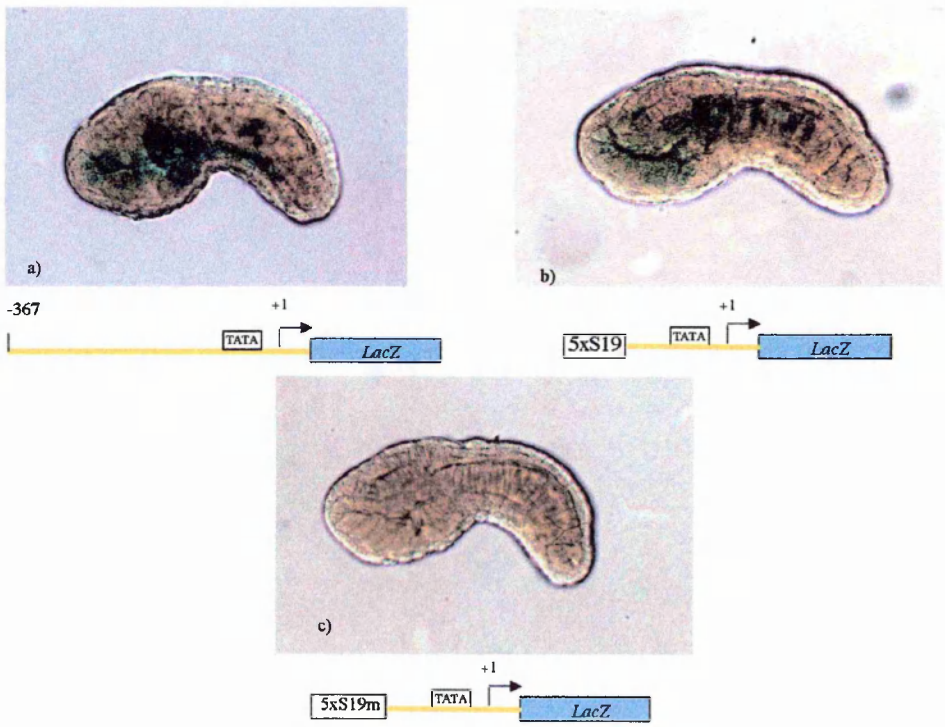


Fig. 23: ***The S19 element is responsible for endoderm activation of Cittf1.*** *Picture b shows a representative embryo electroporated with the plasmid containing 5x S19 oligonucleotide; the mutation m19.5, in 5xS19, abolishes the expression of b-galactosidase (picture c). Also shown are the results obtained in the same experiment by introduction into Ciona embryos of the control construct #5E containing the minimal promoter (picture a).*

fertilized *Ciona intestinalis* eggs via electroporation. After 7 hours, the tailbud embryos were collected, fixed and stained for β-galactosidase activity. Introduction of 5xS19pBSLacZ resulted in the development of embryos that express β-gal activity in endoderm cells of the trunk (fig. 23). The same result was observed in control embryos that received (#5E). Those embryos electroporated with the 5xS19mpBSLacZ (M19.5 mutation) did show any β-galactosidase activity.

This result confirms that R2 site contains a regulatory element (GCATT) that is necessary and also sufficient for transcriptional activation in endoderm lineage of Cititf1 gene. The b-gal signal is also detected in cells of the notochord suggesting that there is a regulatory region that represses Cititf1 expression in these cells.

4.3.7 A new reporter to test regulatory regions in Ciona Intestinalis

When we used the Cititf1 minimal promoter to drive expression of the reporter we observed a very weak signal. Therefore, we decided to use an alternative minimal promoter that would show a stronger expression to test the activity of the element we defined. To this end, we tested the Epstein Barr virus (E1BTATA) minimal promoter. The EIBTATA box was cloned downstream the 5xS19 element (as described in materials and methods and figure 24). The resulting constructs were electroporated into *Ciona intestinalis* fertilized eggs. 7 hours after electroporation, the tailbud embryos were collected, fixed and stained for β-galactosidase activity. The b-gal activity was detected in endoderm cells of the trunk, in

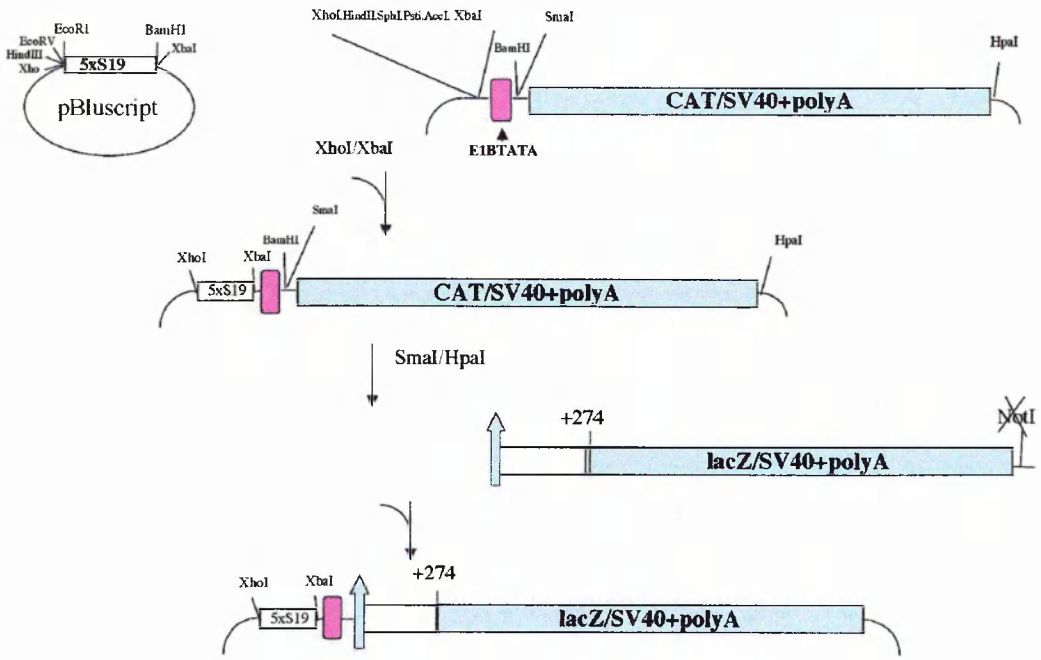


Fig. 24: ***Schematic representation of the strategy to prepare the construct containing five copies of S19 oligonucleotide upstream the E1b TATA box.*** *The 5xS19 inform the pBS plasmid, was cloned in to the XhoI/XbaI sites of the E1B/Cat plasmid . Subsequently, the CAT reporter gene was substituted with the fragment containing the transcription start site of Cititf cloned upstream to LacZ reporter gene.*

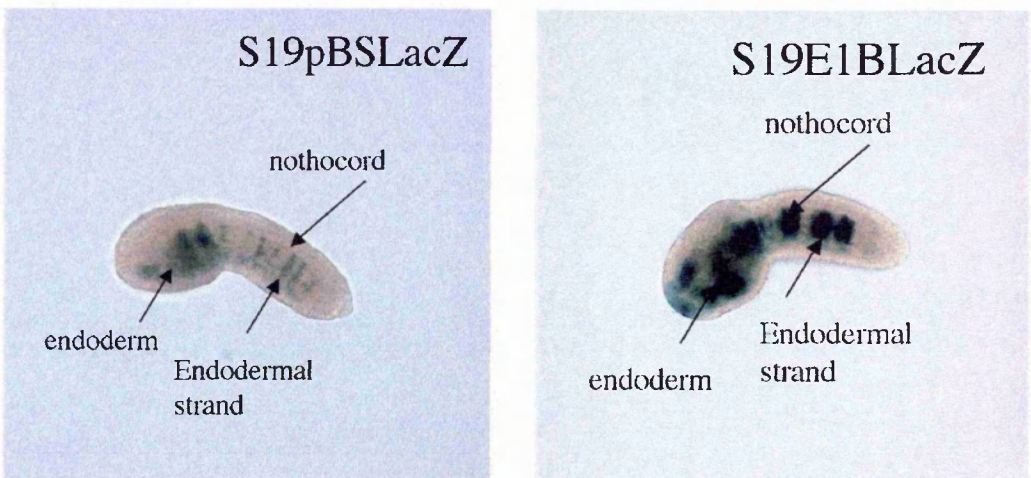


Fig. 25: ***The E1BTATA box enhances the transcriptional activity of five copies of S19 double-stranded oligonucleotide cloned upstream minimal promoter of Cittf1.*** *The picture shows representative embryos electroporated with plasmid 5xS19E1bTATApBSLacZ and 5xS19pBSLacZ, respectively.*

endodermal strand, in notochord and mesenchyme cells in embryo electroporated with 5xS19E1bLacZ after 30 minutes of incubation (fig. 25). In contrast, the embryo electroporated with 5xS19pBSLacZ showed b-gal activity in the same tissues after 6 hours of incu-

bation. Therefore, the E1BTATA minimal promoter can be used to study the cis-regulatory regions of the Ciittf1 gene.

4.4 Discussion

Cititf1 is the earliest zygotic gene identified so far in ascidians thatshows a restricted expression to cells of the endodermal lineage. Therefore, the analysis of the regulatory elements involved in the control of its tissue-specific expression could be helpful in clarifying the molecular pathway contributing to endoderm determination during ascidian's embryogenesis. A previous study characterized a DNA sequence that extends 355 bp upstream of transcription start site and is able to drive the expression of the reporter gene in the endoderm. This enhancer contains distinct regulatory regions whose balance is responsible for Cititf1 activation in the endodermal cell lineage and for its repression in the mesenchyme tissue. No mutation within this region led to identify an endoderm specific cis- acting element, therefore, the presence of redundant regulatory elements has been postulated. The identification of the repeats of eleven nucleotides (R1, R2) has provided more insight into the endoderm specific activation of the Cititf1 promoter. This sequence is repeated two times (R1 and R2) and is likely able to bind the same protein, as demonstrated by the formation of specific DNA-protein complexes that show the eletcromobility in gel shift assay studies.

The importance of the two repeats was also established in vivo. The two repeats do not have

the same role in the regulation of the Cititf1 promoter. The embryos electroporated with the #5E construct carrying the single M15, which falls in R1 site, did not show any reduction in level of expression of β-galactosidase. In contrast, electroporation of the construct bearing the M15 mutation and the M20 mutation, which falls in the R2, produced embryos showing a strong reduction in β-galactosidase expression in the endoderm. These results suggest a cooperative action between the regulatory protein/s that bind this motif. Furthermore, embryos electroporated with construct r1, carrying eleven base pairs mutated in the R1 site, show b-gal expression in the endoderm with comparable intensity to that obtained using the control plasmid #5E, suggesting a minor role of this element in the regulation of the Cititf1 promoter. The R1 site might increase the transcriptional activity of the R2 element. Unfortunately, it is difficult to prove this hypothesis because small differences in the level of transcription are undetectable using this method. The R2 cis-acting element seems to be sufficient to regulate the endodermal expression of Cititf1. In fact, the electroporation of transgene r2, carrying eleven base pairs mutated in R2 site, results in the development of embryos in which the β-galactosidase activity is completely lost from the endoderm territory. Moreover, within the R2 site only five base pairs are responsible for the expression of β-galactosidase in endoderm cells as demonstrated by electroporation of M19,5 construct. The R2 site has been further characterised by biochemical analysis. Band-shift analysis using the oligonucleotide containing the core sequence R2, has clearly demonstrated that this site specifically binds some proteins present in the nuclear extract derived from larva and 110-cell stage embryos. This binding involves exclusively the nucleotides present in R2 site.

Taken together the in vivo and in vitro data suggest a model where the candidate regulatory factor binds to five nucleotides in the R2 site, GCATT. Other proteins cooperate to stabilise this binding interactin with others nucleotides in that site.

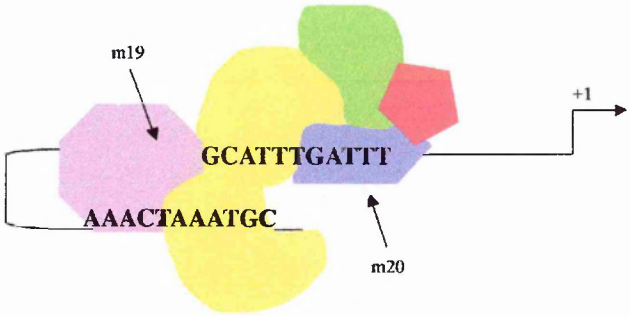


*Fig. 26: **A model representing the formation of transcriptional complexes binding the R2 and R1 site.** The arrows indicates the position of the mutation that strongly reduces the level of expression of the LacZ reporter gene.*

Moreover, the R1 site could reinforce the binding of this complex to the DNA. A schematic representation of this model is shown in the figure 26. The strong reduction in reporter gene expression (LacZ) observed in embryos electroporated with constructs carrying the M19 and M20 mutation, upstream and within R2 site, respectively (Fanelli A et al., 2003), strongly support this model. Furthermore, the transgenic embryo for the plasmid containing five tandem copies of the S19 oligonucleotide cloned upstream to minimal promoter of Cititf1, show b-gal activity in the endoderm of the trunk, in the endodermal strand and also in notochord. This result definitively shows that the R2 site has a fundamental role in activation of Cititf1 in endoderm cells. Furthermore, the expression of the above describe transgene, in ectopic tissue such as mesenchyme and notochord could be due to the absence of regulatory regions necessary for repression of Cititf1 in these tissues. The expression in mesenchyme cells has been already been reported in analysis of the -152 construct (see introduction and Fanelli A at al., 2003). No ectopic expression in notochord cells has been detected so far, there could also be a cis-regulatory signal responsible for repression of Cititf1 in notochord cells. In summary,

Cititf1 expression in the endoderm appears to be controlled by the R2 repeated sequence located in the 300 bp enhancer. The R2 site is necessary and sufficient for Cititf1 activation, and the R1 site cooperates with R2.

This analysis was the stepping-stone to perform the yeast one hybrid screening that allowed to identify the trans-regulatory factors involved in the endodermal expression of Cititf1.

Chapter 5

Result and discussion (iii)

The identification of proteins binding to the R2 site using a yeast one hybrid screen

5.1 Summary

The characterization of the R2 enhancer laid down the basis for a yeast one-hybrid assay, which would allow identifying the transcription factor binding to this element. The multimerized S19 oligonucleotide containing the sequence corresponding to this enhancer and previously characterized by EMSA analysis was used as bait. The oligonucleotide was cloned upstream two reporter genes His 3 and Lacz. His3 encodes the imidazole glycerolphosphate dehydratase enzyme involved in histidine biosynthesis and allows a genetic selection of yeast cells containing it. Lac Z encodes the β-galactosidase enzyme and allows to detect yeast cells by means of a colorimetric assay. The two vectors were integrated separately or together into the genome of a yeast strain that cannot grow on medium lacking histidine. We tested for the presence of natural revertants on these modified strain. Cells containing the two reporter genes were selected as host for a library screen. The cDNA library, to perform the screen, was prepared from RNA derived from an embryo at 110-cell stage. The library was transformed into the yeast strain and the cells were plated on selective medium. About 161 colonies grew in selective media. The plasmids fromthese positive clones was extracted and re-transformed into a tester yeast strain. Only 17 of them confirmed their

ability to confer an his-+ phenotype and β-galactosidase activity. All 17 clones were tested for expression on both 110-cell stage and tailbud stage embryo by in situ hybridization. Only one clone showed expression in endoderm precursor cells in both embryonic stages analysed.

5.2 Introduction

Many eukaryotic transcriptional factors are composed of physically and functionally independent domains: the DNA-binding domain and the activation domain. Thus, it is possible to construct a chimerical protein containing two domains deriving from different transcription factors. This feature has been useful to develop assays in order to isolate DNA-binding proteins or cis regulatory elements. One of these assays was developed for yeast cells. The yeast one hybrid assay allows to identify specific cis regulatory elements recognised by a specific transcription factor, but also to isolate transcription factors that bind to a specific cis regulatory element. To isolate transcription factors, this assay is performed employing a genetic selection system (Chang et al.,1993). Genetic selection for transcription factor genes first requires the creation of yeast strain containing a reporter gene whose expression is controlled by a cis-acting element that binds the transcription factor of interest. Usually, in these screens, two different reporter genes are used, the imidazole glycerol-phosphate dehydratase reporter gene (His3) and β-galactosidase reporter gene (LacZ).

The two reporter genes have different roles in this screening. The presence of a yeast minimal promoter upstream the HIS reporter gene allows this gene to be transcribed at low levels providing a sufficient amount of imidazole glycerol-phosphate dehydratase enzyme, which is essential for histidine biosynthesis. The yeast cells carrying the his3 reporter plasmid are able to grow in a media lacking histidine. However, the low levels of transcription of His3 gene are not sufficient to allow their growth in a media also containing a low concentration of 3-amino-1, 2, 4-triazole (3-AT), an inhibitor of imidazole glycerol-phosphate dehydratase enzyme (fig.27a). However, if the His3 gene is transcribed at a higher level, the yeast cells are able to grow also in presence of low amount of inhibitor. Instead, the β-galactosidase reporter gene allows to perform a colorimetric assay for an easy detection of the yeast cells containing the transcription factor of interest that binds to the cis element cloned upstream of it. The plasmid containing the β-galactosidase reporter also contains an additional marker that is used to select the cells that have integrated this plasmid in their genome. Usually the Ura3 gene marker is used, which encodes orotidine 5' phosphate decarboxilase an enzyme involved in uracile biosynthesis (fig.27 b).

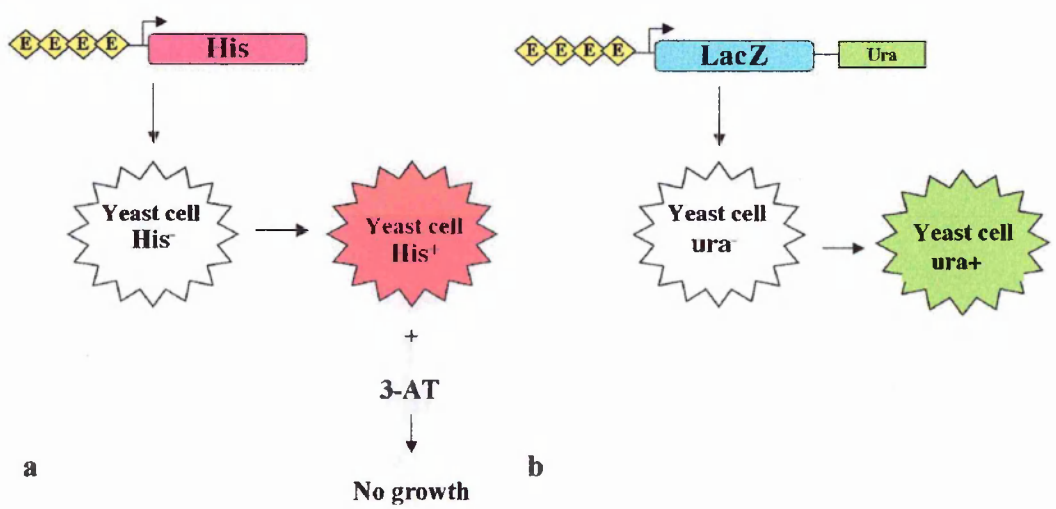


*Fig.27: **Schematic representation of a yeast one hybrid system**. a) His reporter gene b) LacZ reporter gene*

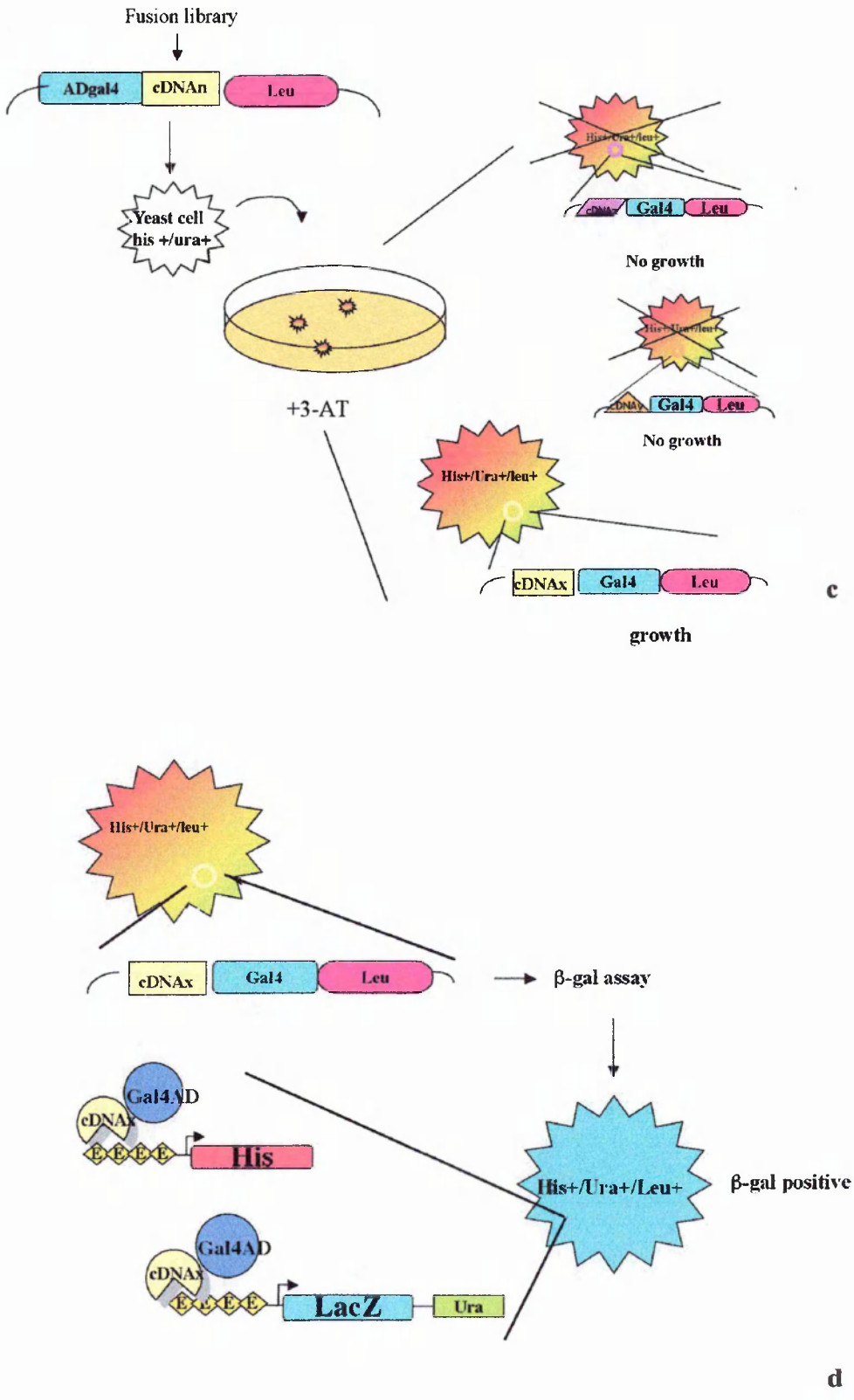


Fig.27: ***Schematic representation of a yeast one hybrid system.*** *c) yeast screen d) isolation of positive clones. See text for explanation.*

This marker allows the yaest strain to grow in absence of uracile. Consequently the yeast cells carrying both reporters can be selected on a medium that lacks histidine and uracile.

Before performing the yeast one hybrid screening these reporter plasmids must be integrated into yeast genome in order to provide high stability and a constant number of gene copies. This operation is carried out by insertion into the his3 or ura3 locus. The cells that have permanently integrated the two reporter plasmids in their genome can be selected by plating on medium lacking the histidine amminoacid and /or the uracile base. This yeast strain can now be transformed with a cDNA library constructed with the mRNA derived from cells that contain the transcription factor of interest. The cDNA library is constructed in a vector that allows the cDNA sequences to be fused to sequence encoding a strong transcriptional activation domain such as Gal4 (See below). Thus, if a sequence encoding a DNA binding domain of a transcrip-

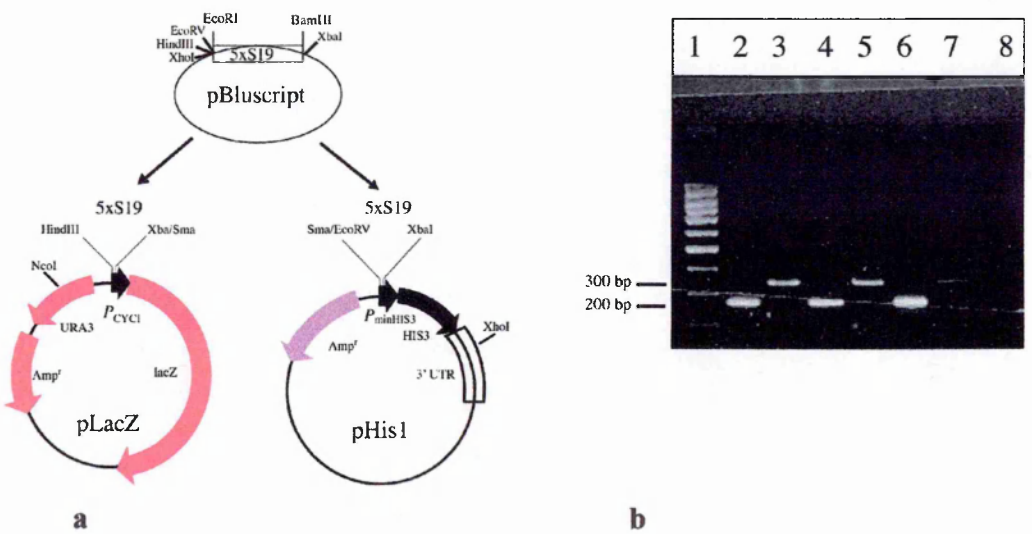


Fig.28 : a) Schematic representation of the cloning strategy; the concatameric oligonucleotide was first cloned into EcoR1 and BamH1 sites of the pBluscriptplasmid in ,subsequently it was cloned into the pLacZ and pHIS1 plasmids as described in the text , b) the insertion into yeast genome of the two linearized plasmid. The S19pHis and S19 pLacZ genomic insertion was tested by PCR on three different clones. ***In lane 2,4 and 6 was loaded the PCR on S19pHis clones respectively 1,2 and 3; in the lane 3, 5 and 7 was loaded the PCR on S19pLacZ clones respectively 1,2 and 3. lane 8 negative control.***

tion factor is expressed as fusion protein, it binds to the cis regulatory region by its DNA-binding domain and activates the transcription of the reporter genes, His3 or LacZ, by the Gal4 activation domain. Upon this activation the yeast cells can grow on selective medium lacking histidine, uracile, leucine[7] and in presence of low concentration of 3-AT (fig. 27c). Moreover, the activation of LacZ reporter produces blue cells that can easily identified (fig. 27d).

5.3 Result

5.3.1 Preparation of the two stable yeast strain YM4271 his+ and ura+, containing five copies of S19 oligonucleotide

In order to increase the probability of protein-DNA interaction, five tandem repeats of the S19 sequence (see material and methods) were inserted into the EcoRI-Xho sites of pHISi1 and the HindIII-SmaI of pLacZ[8] , to generate 5xS19-pHisi1 and 5xS19-pLacZ, respectively (fig. 28a). These two bait constructs were respectively linearized with XhoI and NcoI. The site recognized by these restriction enzymes falls in 3' untranslated region immediately following the HIS3 marker (in pHisi1) or within the URA3 marker (in pLacZ).

The linearized plasmids are separately inserted by chemical transformation in competent

---

[7] *Leu3 is the gene marker selection for plasmid containing the cDNA library. Leu3 encodes 3-isopropylmalate dehydrogenase and is involved in leucine biosynthesis.*

[8] *The pHISi1 and pLacZ plasmid are provided from Clontech Company.*

yeast cells that are his3-, ura3- phenotype and undergo selection on minimal medium lacking histidine and/or uracile. About 30 colonies grew after selection on minimal medium (see material and methods) lacking histidine and more than 100 colonies grew on minimal medium lacking uracile, as expected from the different susceptibility of the integration site for the two genes. Three colonies for each transformant were picked and streaked on the selective minimal medium. The plasmids p53pLacZ and p53pHisi-1, containing the p53-target sequence, were also transformed in yeast strain and used as positive controls. In this case 18 colonies grew after selection on minimal medium lacking of histidine and 200 colonies grew on minimal medium lacking uracile.

These plasmids do not have an origin of replication and are not able to replicate autonomously therefore they are stably integrated in the yeast genome. After selection on the minimal medium, a PCR analysis on DNA deriving from each clone was performed to confirm that the integration process of plasmids in the yeast genome did not affect the integrity of the two transgenes.

To this purpose, PCR oligonucleotides were designed to the vector region that flanks the multicloning site of plasmids pHisi1 and pLacZ. In the case of pHisi1 plasmid, the sequence extending between the two oligonucleotides is 100 base pairs long before insertion of the 5xS19 oligonucleotide and should be 200 bp after insertion of 5xS19. In the case of pLacZ, the primers were chosen such that in the absence of 5xS19, a 200 bp fragment should be amplified. In the presence of 5xS19 a band of 300 base pairs should be amplified. As shown in the figure 26b we obtained three clones for each plasmid. In fact the bands cor-

responding to fragment containing 5xS19 are respectively of 200bp for genomic DNA derived from transformed yeast cells containing His reporter construct and 300bp for genomic DNA derived from transformed yeast cells containing LacZ reporter construct. This data confirm that the genome of yeast reporter strains contains five copies of S19 oligonucleotide.

5.3.2 Evaluation of revertants in yeast transformed strains and preparation of the double reporter strain YM4271S19His/LacZ

After the preparation of the two single reporter strains, the frequency of spontaneous auxotrophic revertant was analysed. Independent clones, from each transformed yeast strain, were tested on media lacking histidine amminoacid or uracile base. In addition, the clones carrying the His reporter gene (YM4271S19pHis strain) were transformed with the S19pLacZ reporter construct while, the clones having the LacZ reporter gene (YM4271S19pLacZ strain) were transformed S19pHis reporter construct in order to make a dual reporter strain. The yeast cells transformed with the two reporter constructs were plated on minimal medium lacking histidine amminoacid and uracile base. We selected those cells, which containing both selective markers, grew on the above-mentioned minimal medium. These rèsults are summarized in Table 1. Strain YM4271(his-ura-) transformed with S19pHis grows on media without histidine as well as on complete medium (YPD), and can not grow without uracile. When the modified yeast strain is transformed with S19pLacZ plasmid that contains the URA marker, few colonies grow.

STRAIN	PLASMID	MEDIA SELECTION	CFU
YM4271 S19pHis(cl.1)	H_2O	YPD	2000
YM4271S19pHis(cl.1)	H_2O	SD/-U	0
YM4271S19pHis(cl.1)	H_2O	SD/-H	2300
YM4271S19pHis(cl.1)	S19pLacZ 10μg	SD/-U/-H	12
YM4271S19pLacZ(cl.2)	H_2O	YPD	1000
YM4271S19pLacZ(cl.2)	H_2O	SD/-U	2150
YM4271S19pLacZ(cl.2)	H_2O	SD/-H	0
YM4271S19pLacZ(cl.2)	S19pHis 10μg	SD/-U/-H	44

Table1. Test to identify natural auxotrophic revertants of the yeast strains. See text for explanation

The table also shows that also in the case of the yeast strain containing the S19pLacZ, no background growth was observed. However, more colonies grew when this strain was transformed with S19pHis. These results suggest that the frequency of natural auxotrophic revertant was equal to 0 both cases.

Several colonies, obtained after transformation of YM4271S19pHis with S19pLacz reporter construct were picked and streaked again on selective minimal medium. One colony was used to prepare the competent yeast dual strain to screen the fusion library Gal4/110cDNA (see below).

During selection, the basal level of HIS expression could result in the growth of colonies that

amounts of 3-AT to the minimal medium. As the concentration of 3-AT increases the size of colonies decreases until no growth is observed. The choice of the optimal concentration depends on a balance between the size of the colonies, that must be very small, and a low concentration of inhibitor. In the case of the dual reporter strain YM4271pHis/lacZ the optimal concentration of 3-AT was estimated by plating the cells on minimal medium containing 15, 30, 45 and 60mM of 3-AT. We determined the best concentration to be 30mM. This concentration allows the yeast cells to grow only at high levels of the his3 gene transcription.

5.3.3 One hybrid screen to identify R2 binding proteins

The expression of Cititf1 begins to be detected in the endoderm precursor cells at 110-cell stage, thus the factor activating its expression should be already present at this stage. To isolate this factor a cDNA library was prepared with total RNA extracted from 110-cell stage embryos. This library was cloned in the yeast expression vector pGAD424 (fig.29) In order to increase the copy number of cDNA molecules the library was transformed in DH5a electrocompetent bacterial cells. The transformants were plated on LB media and about $0.7x10^6$ cfu grew. After amplification, the DNA was extracted from bacterial cells (see material and methods).

The yeast vector pGAD424 contains the marker LEU3, which allows to control the number of yeast cells transformed. To establish how many different cDNAs molecules must be tested in the yeast one hybrid screen, a small-scale transformation was performed. 1,5 micrograms of the empty vector pGAD424 and the same amount of DNA derived from the library, were transformed

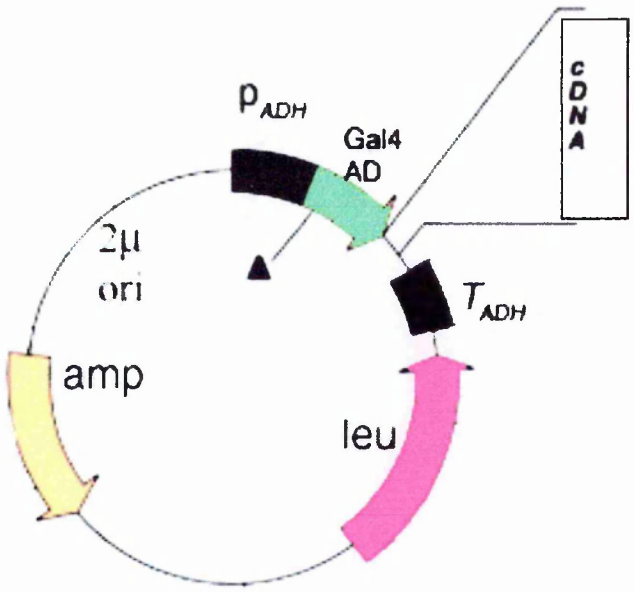


Fig. 29 : ***Map of pGAD424****. pGAD424 encodes the activation domain of the yeast GAL4 transcriptional activator .The cDNA lybrary was cloned in EcoRI site in frame with Gal4 activation domain. The AD is expressed at low levels in yeast host cells from a truncated ADH1 promoter and is targeted to yeast nucleus by SV40 T-antigen nuclear localization sequence*

in the dual reporter yeast strain YM4271S19lacZ/his. An aliquot of this transformation was plated on media either lacking leucine or lacking of histidine, uracile, leucine and in presence of 3-AT. Of $3x10^4$ independent clones tested, only 13 colonies grew when the dual reporter yeast strain was transformed with fusion library, no growth was observed with the empty vector.

Therefore, to perform the large-scale screen 20 micrograms of DNA of the fusion library were transformed in the dual reporter strain. In this experiment the total number of independent clones tested on minimal medium lacking leucine was about of $0.5x10^6$ while the number of colonies that grew on minimal medium containing 30mM 3-AT but lacking histidine, uracile and leucine was 161.

To eliminate the false positive clones four different tests were performed: (i) growth on his-/ura-/leu- selective medium containing 3-AT, (ii) elimination of false positive by reselection on the same selective medium and PCR screening to eliminate colonies with the empty vector, (iii)

extraction of the plasmids and re-transformation in the dual reporter strain, and (iv) analysis of b-galactosidase activity. All 161 colonies were streaked on the minimal medium and only 133 colonies grew. To further eliminate the false positives, these colonies were screened by PCR. 50% of the clones had an insert and the average size was about 900 bp. All 58 independent clones were cultured in minimal medium lackingleucine and the plasmids, containing the cDNA encoding the candidates R2-target binding proteins, were extracted and transformed in E. coli cell (Top 10). In some case the PCR analysis showed the presence of two different clones in the same yeast cell, therefore five colonies for each transformation were sequenced at the 5' and 3' end. In some cases the presence of two different plasmids in the same yeast cell was confirmed. Although more than 400 bp were sequenced from both ends, in most cases the complete sequence of each clone is not available. All clones were compared and aligned. Thirty-six different sequences were compared to nucleotide and protein databases, the *Ciona intestinalis* database (TIGR, UNIGENE) and to protein database (swiss-prot, InerPro and Prodom). Protein identity was considered only for an homology better than 30% on a stretch of 40 amminoacids.

Subsequently, in order to eliminate the false positives, the 36 plasmids extracted from E.coli, were transformed again in the YM4271S19lacZ/His strain and plated on selective minimal medium (SD-his/-ura/-leu/+3AT). 26 clones grew after 6 days. A b-galactosidase assay was performd on the filter. The manufacture's instruction states that blue colonies should be detected within 8 hours of incubation. We observed blue staining, albeit with different intensity, in 17 clones and our positive control (Gal4/p53 fusion protein and p53-target yeast reporter strain), 10 hours after incubation. The results are also shown in table 2.

n° of clone in the cluster	Lenght of contig to 5' end	blast N-Ciona db	hit-blastX	E-value	b-gal positive
group of 18 clones	1196	citb028m05	NADH dehydrogenase subunit 3 *Ciona intestinalis*	0	
131a-26	478	cilv019a20	Ciprincle (*Ciona intestinalis*)	0	+
157-20a-62	506	cieg041d22	TBP-like protein (*Zea Mais*)	9.00E-18	+++
158-15	705	ciad069n06	no hit		+++
2	851	citb047h22	H-CAP -H protein (*Homo Sapiens*)	5.00E-18	
3	976	cign069g08	no hit		+/-
7	916	ciad009i08	oxysterol-binding protein (*Mus Musculus*)	2.00E-65	+++
20	400	citb049m10	Hypotetical protein (*Homo Sapiens*)	2.00E-22	
24	381	cicl016d22	cathepsin B precursor	2.00E-87	
50	597	cibd081n02	Zinc nuclear DNA-binding motif	1,00E-53	++
72	967	ciad091e06	TRPO (*Homo Sapiens*)	5.00E-31	+
104	358	cieg56n06	no hit		+
106	54	ciad047n22	ST-13like tumor suppressor (*Homo Sapiens*)	2.00E-40	
110	940	cibd042o06	hypotetical protein (*Homo Sapiens*)	0	+
118	500	citb068l06	ubiquitine conjugate enzyme (*Homo Sapiens*)	0	+
119	402	cicl027k03	microsomal glutation-S-transferase (*Homo Sapiens*)	1.00E-28	+
121	441	ciad042f22	ribosomal protein S3 (*Homo Sapiens*)	1.00E-110	
123	484	cieg023m02	no hit		+
132	522	cicl016h02	rhomboid-like associated presenilins (*Homo Sapiens*)	4.00E-34	+
138	686	citb089b12	no hit		+
144	575	cigd034e04	splicing factor SPF30 (*Homo Sapiens*)	4.00E-31	+
148	516	cieg80c22	Hypothetical protein F35F10.12 protein *C. elegans*	1,00E-21	
152	529	cibd031e05	Isocitrate dehydrogenase [NAD] subunit beta (*Homo Sapiens*)	1.00E-96	
159a	549	cinc004c20	NcK adaptor protein (*Homo Sapiens*)	1.00E-38	+
159c	421	cicl031l23	bromodomain PHD finger transcription factor (*Homo sapiens*)	5,00E-40	+++

Table 2. Clones isolated after the yeast one hybrid screening

5.3.4 In situ hybridization analysis

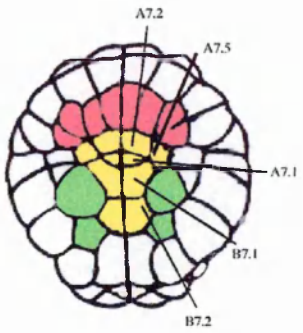
The sequence alignment did not help in identifying the putative proteins that bind the R2 endoderm specific element. To establish if any of these clones are expressed in endoderm precursor cells, I decided to perform an in situ hybridization screening. Only the clones that were positive for b-gal assay were chosen as template for riboprobe synthesis.

The plasmid in which the cDNA library was made does not allow preparing antisense riboprobes. Therfore, to produce the RNA probes, the insert of each cDNA plasmid was amplified by PCR and subcloned into TA-topo cloning vector containing two different promoters T7 and Sp6. All clones were sequenced verified. The in situ hybridization was performed on 110-cell and tailbud stage embryos. Only three clones seem to have an interesting expression pattern, as show in the fig. 30

The gene represented by clone 50 is expressed in all endoderm precursor cells at 110-cell and tailbud stage embryos. The transcript is also detected in notochord and mesenchyme precursor cell in both embryonic stage analysed. A database search showed that the putative ORF of clone 50 encodes a protein containing a protein-protein interaction domain and zinc finger domain found in c-cbl oncogene.

Clone 131 shows high identity with Ci-prinkle, which encodes a LIM-containing protein, already identified during a screening of Ci-Brachyury-downstream target gene (Hotta K. et al., 2000). The Lim domain is a zinc finger domain responsible for protein-protein interactions. The transcript of this gene has been detected in notochord cells in embryos at the tail-

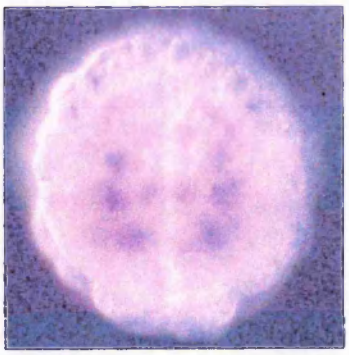
a



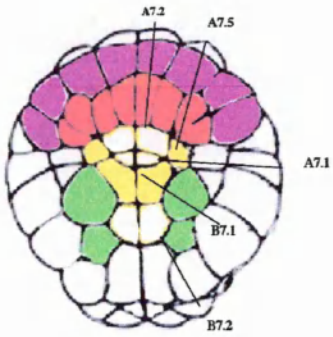
CLONE 50

b

A



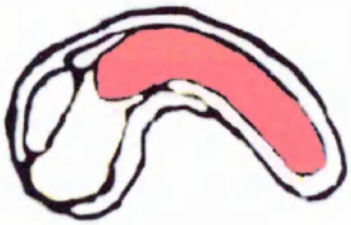
a'



CLONE 131

b'

B



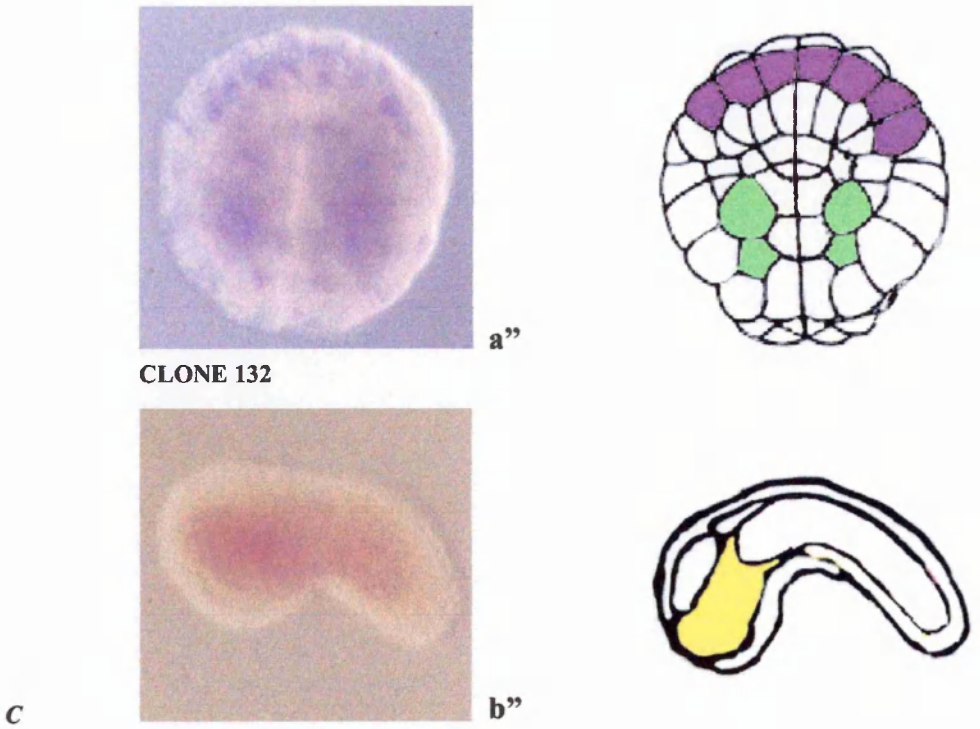


Fig.30 : ***In situ hybridization of clones 50,131 and 132.*** *Whole-mount embryos at 110-cell stage (a-a") and at tail-bud stage (b-b"), were hybridized with digoxigenin labeled antisense RNA probes, and the signal was detected by histochemical staining with alkaline phosphatase. The embryos are oriented to display the vegetal view in a) and lateral view in b).*
A) *The transcript of clone 50 is detected in endoderm, mesenchyme and nothocord cells in both 110-cell and at tailbud stage embryos;* ***B)*** *the transcript of the clone 131 is detected in mesenchyme, in B7.1, in A7.5 blastomeres in 110-cell stage embryos and in notochord cells in embryos at tailbud stage;* ***C)*** *the transcript of 132 clone is detected in nervous system precursor cells and in mesenchyme in 110-cell stage embryos and in endoderm of the trunk in embryos at tailbud stage. Cartoons indicating 110-cell stage and tailbud stage are also shown; the cell marked after in situ are stained and the color correspond to specific tissue (see fig. 1)*

bud stage and in blastomeres at the 110-cell stage, which are respectively precursors of endoderm and mesenchyme cells.

The gene of clone 132 is mainly detected in notochord precursor cells but a weak signal is also detected in the mesenchyme cells at 110-cell stage embryos. The endoderm of the trunk seems is only labeles at the tailbud stage. A database search showed that the putative ORF of this clone encodes a protein with high similarity to membrane rhomboid-like proteins involved in dorso-ventral patterning of Drosophila embryo

5.4 Discussion

In order to identify DNA-binding proteins that regulate the expression of the Ci-titf1 gene we conducted a yeast one hybrid screen. The optimal conditions to perform this assay were determined first to avoid a large number of false positives. The plasmids containing the five copies of the S19 oligonucleotide driving expression of the reporter genes were integrated into the yeast genome to produce cells with a high number and stable copies of the plasmids. These cells were selected on minimal medium and the possible presence of natural revertants was also investigated. The yeast reporter strain YM4271S19lacZ/his, which did not show the presence of natural revertant, and that contains the two reporter genes was used to screen the Gal4/cDNA-110 fusion library. Of 161 clones identified as positive, 26 were able to induce the activation of histidine reporter gene after transformation in the reporter strain. Only subset of these (17) was also able to activate the expression of the b-galactosidase reporter gene. Some of these clones show homology to known proteins, such as DNA binding proteins characterized by zinc finger domains.

The expression pattern of these 17 was investigated by in situ hybridization. This type of analysis allowed reducing the number of putative candidates that could play a role in activation of Cititf1. In fact, only three clones were detected in endoderm precursor cells. The endodermic expression pattern was assessed in both the 110-cell and tailbud stage embryos. The transcript of clone 131 was only detected in the endoderm at the 110-cell stage embryo; clone 132 is detected in the endoderm at the tail-bud stage. This suggests that these two genes (131-132) are involved in endoderm differentiation but do not regulate Cititf1 activa-

tion that is expressed at both stages. The transcript of the clone 50 is detected in the endoderm precursor cell at both stages analysed. Its expression is also detected in notochord and mesenchyme precursor cells. The putative coding sequence of clone 50 shows high identity with the protein-protein interaction domain and the zinc-finger domain of the human protein c-Cbl. Although the primary structure of the c-Cbl protein resembles that of a DNA-binding transcription factor including a basic region, a nuclear localization sequence, a zinc finger-like motif and a leucine zipper, its exact role in the cell is still not fully elucidated. Some data on its putative transcriptional activity comes from studies on its nuclear localization (Denis G et al., 1999) and from the analysis of its expression in differentiating and proliferating cells. Differentiation of both erythroleukemia cells and teratocarcinoma cells show a decrease in c-cbl expression, with kinetics similar to those of transcription factors that are immediate early response genes. No change in expression is observed in proliferating fibroblasts and spleen cells. The truncated form of c-cbl, as v-cbl, enters into nucleus and binds DNA, contributing to neoplastic transformation of B lymphocytes and fibroblasts (Mushinski JF, et al. 1994).

Recent reports have identified other cbl-like proteins involved in the negative regulation of receptors or non-receptor tyrosine kinases through the ubiquitin pathway. Hakai, a cbl-like protein recently identified, has been suggested to be involved in the regulation of epithelial/mesenchymal transitions in development and metastasis (Fujita Y, et al., 2002). Hakai contains SH2, RING, zinc-finger and proline-rich domains, and interacts with E-cadherin inducing ubiquitination of the E-cadherin complex. Expression of Hakai in epithelial cells

disrupts cell-cell contacts and enhances endocytosis of E-cadherin and cell motility. The RING-finger domain in c-cbl, has been demonstrated to be involved in recruitment of ubiquitin-enzyme E3 (Yokouchi M et al., 1999).

The sequence of clone 50 is incomplete so we can not exclude that this protein contains other domains, other than the protein–protein interaction domain and a zinc-finger domain. We could not identify a site for ubiquitin modification, therefore, it is reasonable suppose that clone 50 is not involved in the ubiquitation process. The zinc finger domain is present in many proteins and is responsible for binding to nucleic acids and/or to other proteins. The association of many zinc finger proteins with DNA- and/or protein- binding domains allows the formation of multiprotein complexes. Within these complexes, the DNA-binding motifs recognize a target sequence in a specific manner and the protein-protein interaction domain allow to assemble multiprotein regulatory complexes commonly involved in chromatin remodelling (Bach et al., 1999) and in transcriptional regulation.

Clone 50 could regulate endoderm differentiation by interacting with other protein through its SH3 domain and activating the Cititf1 promoter by its zinc finger domain (fig. 31). Two considerations suggest that this gene could be involved in the endoderm differentiation through this mechanism. The first one is that the clone 50 was isolated by means of the yeast one hybrid assay, using as a target sequence the R2 endoderm specific element of the Cititf1 promoter, suggesting that this protein can bind DNA. The second is that the expression pattern of clone 50 cDNA overlaps the expression pattern of the transgenic 5xS19pLacZ. The multimerized S19 oligonucleotide is able to drive the expression of LacZ

MASGAMNGNTAQQRHAQRQGPIRILNHFRGFSDALNDVTPHHQRVTIDK
KYIEKSSKLMDKVAKFCQQSKMNLKNSPPYILDILPDTYNHLRLIMLRY
EDRLQVLNECEYFRIFLDNLIKKCKEAIKLFREGRERMFDEN
SQQRRSLTKLSLVFSHMLAELKAEFPDGRFIGEDYRITKSEAAN
FWKTTFGTRTIVQWKIFRQELNKVHPISSGLECIALKSTIDLVCNNYI
SNFEFDVFTRLFQPWSNLLKNWNMLAVTHP**GYMAFLTYDEVKAR**
LLKYIHKPGSYIFRLSCTRLGQWAVGYVTMDGQILQTIPQNKSLCQA
LIDGWKEGFYLFPDGRQVNPDLSGLLEQSEDDEEHITVSQEQYELYCEM
GSTFQL KI AENNKDVRIEP GHLI KS LESWQEMDNSA
SPT PW RCEIKGTETVIIEPYESEKKEEEEATGDKTCKKEKEEEKKEEN
KEEEDLFGEIGVSVNKLPVSSCCSNKNILGSMLINCC

*Fig.31 : **Putative coding sequence of clone 50.** The putative coding sequence of clone fifthy contains an SH2 domain responsible for protein-protein interaction that is underlined and a zinc finger domain indicates by C (in yellow). The spacing of the cysteines in this domain is C-x(2)-C-x(9 to 39)-C-x(1 to 3)-H-x(2 to 3)-C-x(2)-C-x(4 to 48)-C-x(2)-C.*

reporter gene in endoderm of the trunk but also in notochord cells of a tailbud embryo; the clone 50 is expressed in notochord cells and endoderm of the trunk at tailbud stage.

However, to confirm that this protein binds the Cititf1 promoter additonal bind-shift analysis should be performed using the in vitro translated protein, the S19 oligonucleotide as probe and several oligonucleotides carrying mutations in R2 site. In vivo experiments including over-expression and /or mis-expression of this protein in Ascidian embryos could clearly confirm that it acts upstream to Cititf1.

The expression of this protein is also detected in precursor cell of tissue as notochord and mesenchyme, therefore, a putative repression site should be also identified in Cititf1 promoter.

Chapter 6

General discussion

Ascidians are marine protochordates that have been used as a model for embryological studies for over a century. In the last decade this organism has also become popular for developmental biology studies. In their larval anatomy, they clearly show some features typical of chordates such as notochord, gill slits, tail muscle, patterned and dorsal CNS. The egg develops into tadpole larvae 16 hours after fertilization. Its embryogenesis has been described in detail and it is possible to predict which cells of the early embryo give rise to which cells of the tadpole larvae. Ascidan embryos provide a unique model with which to study the molecular mechanisms of fate determination in all cell types because the tadpole consists of a small number of cells and few types of tissue.

The identification of factors involved in endoderm commitment has been the goal of this thesis. In contrast to mesenchyme and ectoderm differentiation, endoderm differentiation is an autonomous process that is independent of cell-cell interaction. During the first steps of endoderm differentiation, the combination of several transcription factors leads to the activation of the endoderm differentiation pathway. The small genome of the ascidian and the relative ease to generate transgenic embryos make the ascidian an excellent model to study cis regulatory regions and to analyse the gene function by misexpression or overexpression.

Cititf1 is the earliest zygotic gene identified so far in ascidians that shows a restricted expres-

sion to cells of the endoderm lineage. In contrast to the vertebrate homologue that is thought to regulate the differentiation of gut associated organs as thyroid and lung, Cititf1 is involved in early events of endoderm differentiation in *Ciona intestinalis* as demonstrated by overexpression, interference and misexpression experiments (Ristoratore F et al., 1999, Spagnuolo A et al., 2002). These experiments have demonstrated the capability of TTF1 to alter the differentiation of notochord to endodermal fate, probably recruiting notochord blastomeres. The identification of factors involved in the regulation of Cititf1 could provide more information on the genetic cascade controlling endoderm specification. The analysis of the promoter region of Cititf1, taking advantage of the powerful electroporation method, allowed to identify endoderm specific cis-regulatory elements and to develop a yeast one-hybrid assay to identify the protein/s binding to this element.

In detail, two relevant sites for Cititf1 endoderm expression were identified, R1 and R2, each a repeat of eleven nucleotides. The EMSA analysis showed that these sites could be recognised by the same protein. Surprisingly, the two sites do not have the same importance; the R2 element has a more important role in the activation of Cititf1, than the R1. Probably the protein that recognises and binds these sites does so through increasing affinity. It is also possible that the formation of a dimer makes the complex more stable to activate the promoter.

Interestingly, analysis for putative binding sites has revealed that the mutation introduced in m19.5, located at the most 5' end of the R2 site, covered a potential recognition site for Oct-1, a transcription factor that contains a POU homeodomain. This observation is very interesting because Oct-1 family members have been implicated in the transcriptional regulation of the mammalian TiTf1 (Bingle CD et al., 1996; Ikeda K et al., 1996). Moreover, another member of the

Oct family, Oct-4, has been shown to be essential for endoderm specification in zebrafish (Lunde et al., 2004; Reim G et al., 2004) and the mouse orthologue is involved in differentiation of stem cells in endoderm tissue (Guo Y et al.,2002).

The 3'-most part of R2 site contains the consensus binding site for transcription factors belonging to SRY family and the mutation M20 that falls in this consensus is responsible for a reduction in the expression level of b-galactosidase. The fact that proteins from the HMG super family, such as Sox17, Casanova and Tcf/Lef, have been shown to play important roles in endoderm specification of different organisms is relevant to the tissue specific activation of Cititf1. Furthermore, the SRY-related HMG-box factor, Sox17, has been shown to be a key component for endoderm determination process in Xenopus (Hudson C et al., 1998), zebrafish (Alexander J et al., 1999) and mouse (Kanai-Azuma M et al., 2002).

The Oct and Sox transcription factors appear to form a complex that binds to juxtaposed sites within the enhancer element of several genes and mediates their activation (Ambrosetti DC et al., 1997; Di Rocco G et al., 2001; Guo Y et al., 2002; Tomioka M et al.,2002; Avilion AA et al., 2003; Lunde et al., 2004). Interestingly, some of the genes regulated by this complex are involved in endoderm differentiation (Guo Y et al., 2002; Avilion AA et al., 2003; Lunde et al., 2004). Although the mutagenic analysis of the Cititf1 promoter suggest that these proteins could be involved in its activation, these transcription factors were not identified in the yeast one hybrid screening.

All clones obtained by the yeast one-hybrid screening have been selected for the ability to bind the R2 element. However, many of these clones are false positives and could be due to a non-spe-

cific binding of these proteins to R2. Ideally, only one -(or few) of the proteins represented in the positive clones should be capable to specifically bind to this element and these should be represented in many clones. This was not the case in the screen with the R2 element, in fact, several factors were obtained but each one is represented in few clones. However, the short half-life and rapid turn-over of the mRNA encoding transcription factors involved in differentiation process could explain this result, and help understand the failure to isolate members of the Oct or Sox family. The DNA binding proteins isolated by this screening are characterized by a zinc-finger domain. Recently, a LIM-homeodomain containing protein, Ci-lhx, has been demonstrated to be involved in endoderm differentiation of *Ciona intestinalis* (Satou Y et al., 2001). Interestingly, this gene is expressed in the same cells as clone 50 (notochord, endoderm and mesenchyme), identified in our screen. It is possible that the two proteins interact and activate Cititf1, although we can not exclude the possibility that clone 50 interacts with a transcription factor belonging to the Sox family. Interstingly, the murine Sox2 is involved in differentiation of stem cells in the endoderm (Avillion AA et al., 2003). To confirm these different possibilities, it is important to establish if the ascidian's counterpart of murine Sox2, identified by genome sequencing (Yamada L et al.,2003), is expressed in endoderm precursor cells and its role in activation of Cititf1.

The role of Clone 50 in the activation of Cititf1 could be studied by misexpression using a notochord specific promoter, such as Brachyury's promoter. If Clone 50 regulates Citif, its ectopic expression would induce the same phenotype observed when Cititf1 is misexpressed (Spagnuolo A et al., 2002) and activate the expression of Cititf1 in notochord cells. This type of experiment would clearly demonstrate the role of clone fifty in the activation of Cititf1.

Reference

Alexander, J. and Stainier, D.Y.R., (1999). A molecular pathway leading to endoderm formation in zebrafish. *Current Biology* 9, 1147-1157

Alexander, J., Rothemberg, M., Henry, G.L., Stainier, D.Y.R., (1999) Casanova plays an early and essential role in endoderm formation in zebrafish. *Developmental Biology* 215, 343-357

Ambrosetti DC, Basilico C, Dailey L. (1997) Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Molecular Cell Biology* (11):6321-9.

Asashima M, Nakano H, Uchiyama H, Sugino H, Nakamura T, Eto Y, Ejima D, Nishimatsu S, Ueno N, Kinoshita K. (1991) Presence of activin (erythroid differentiation factor) in unfertilized eggs and blastulae of Xenopus laevis. Proceeding of the National Academy of Sciences of the United States of America 88 (15): 6511-4.

Avilion AA, Nicolis SK, Pevny LH, Perez L, Vivian N, Lovell-Badge R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes and Development.* 17(1):126-40.

Bach I, Rodriguez-Esteban C, Carriere C, Bhushan A, Krones A, Rose DW, Glass CK, Andersen B, Izpisua Belmonte JC, Rosenfeld MG. (1999) RLIM inhibits functional activity of LIM homeodomain transcription factors via recruitment of the histone deacetylase complex. *Nature Genetics* 22(4):394-9.

Barker N, Morin PJ, Clevers H. (2000) The Yin-Yang of TCF/beta-catenin signaling. *Advance in Cancer Research*; 77:1-24. Review.

Bingle CD, Gowan S. (1996) Oct-1 interacts with conserved motifs in the human thyroid transcription factor 1 gene minimal promoter. *Biochemestry Journal* 319 (Pt 3):669-74.

Bowles, J., Schepers, G. and Koopman, P., (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Developmental Biology.* 227, 239-255

Casey ES, Tada M, Fairclough L, Wylie CC, Heasman J, Smith JC (1999) Bix4 is activated directly by VegT and mediates endoderm formation in Xenopus development. *Development.* 126(19):4193-200

Chan JY, Han XL, Kan YW. (1993) Cloning of Nrf1, an NF-E2-related transcription factor, by genetic selection in yeast. *Proceeding of the National Academy of Sciences of the United States of America* 90(23): 11371-5

Chen Y, Schier AF (2001) The zebrafish Nodal signal Squint functions as a morphogen. *Nature.* 411(6837): 607-10

Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate- phenol-chloroform extraction. *Analytical Biochemistry* 162, 156-159

Clements D, Friday RV, Woodland HR. (1999) Mode of action of VegT in mesoderm and endoderm formation. *Development* 126 (21) 4903-11

Clements D, Rex M, Woodland HR. (2001) Initiation and early patterning of the endoderm. *International Review of Cytology.* 203: 383-446

Clements D, Woodland HR. (2003) VegT induces endoderm by a self-limiting mechanism and by changing the competence of cells to respond to TGF-beta signals. *Developmental Biology* 258(2): 454-63

Conlon FL, Lyons KM, Takaesu N, Barth KS, Kispert A, Herrmann B, Robertson EJ. (1994) A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development.* 120(7): 1919-28.

Corbo, J. C., Levine, M., and Zeller, R. W. (1997) Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, Ciona intestinalis. *Development* 124, 589-602.

D'Ambrosio P, Fanelli A, Pischetola M, Spagnuolo A. (2003) Ci-GATAa, a GATA-class gene from the ascidian Ciona intestinalis: isolation and developmental expression. *Developmental Dynamics* 226(1):145-8.

David NB, Rosa FM. (2001) Cell autonomous commitment to an endodermal fate and behaviour by activation of Nodal signalling. *Development.* 128(20): 3937-47

Davidson EH, Rast JP, Oliveri P, Ransick A, Calestani C, Yuh CH, Minokawa T, Amore G, Hinman V, Arenas-Mena C, Otim O, Brown CT, Livi CB, Lee PY, Revilla R, Schilstra MJ, Clarke PJ, Rust AG, Pan Z, Arnone MI, Rowen L, Cameron RA, McClay DR, Hood L, Bolouri H. (2002) A provisional regulatory gene network for specification of endomesoderm in the sea urchin embryo. *Developmental Biology.* 246(1): 162-90.

Denis G, Mandard S, Humblet C, Verlaet M, Boniver J, Stehelin D, Defresne MP, Regnier D. (1999) Nuclear localization of a new c-cbl related protein, CARP 90, during in vivo thymic apoptosis in mice. *Cell Death and Differentiation.* 6(7):689-97.

Denny, P., Swift, S., Brand, N., Dabhade, N., Barton. P. and Ashworth A., (1992), A conserved family of genes related to the testis determining gene, SRY. *Nucleic acid research.*, 20, 28-87

Di Rocco G, Gavalas A, Popperl H, Krumlauf R, Mavilio F, Zappavigna V.(2001) The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function. *Journal of Biological Chemistry.* 276(23): 20506-15.

Dickmeis T, Mourrain P, Saint-Etweenne L, Fischer N, Aanstad P, Clark M, Strahle U, Rosa F. (2001) A crucial component of the endoderm formation pathway, CASANOVA, is encoded by a novel sox-related gene.*Genes and Development.* 15(12): 1487-92

Ding J, Yang L, Yan YT, Chen A, Desai N, Wynshaw-Boris A, Shen MM. (1998) Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. *Nature.* 395 (6703): 702-7.

Dohrmann CE, Kessler DS, Melton DA. (1996) Induction of axial mesoderm by zDVR-1, the zebrafish orthologue of Xenopus Vg1. *Developmental Biology* 175(1): 108-17

Engleka MJ, Craig EJ, Kessler DS (2001) VegT activation of Sox17 at the midblastula transition alters the response to nodal signals in the vegetal endoderm domain. *Developmental Biology.* 237(1):159-72.

Ericson L.E. and Fredriksson (1990) Phylogeny and Ontogeny of the thyroid gland. The thyroid Gland ,Raven Press, Ltd., New York

Erives A, Levine M (2000) Characterization of a maternal T-Box gene in Ciona intestinalis. *Developmental Biology* 225(1):169-78

Fanelli A, Lania G, Spagnuolo A, Di Lauro R. (2003) Interplay of negative and positive signals controls endoderm-specific expression of the ascidian Cititf1 gene promoter. *Developmental Biology.* 263(1):12-23.

Feldman B, Gates MA, Egan ES, Dougan ST, Rennebeck G, Sirotkin HI, Schier AF, Talbot WS., (1998) Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* 395 (6698): 181-5

Fujita Y, Krause G, Scheffner M, Zechner D, Leddy HE, Behrens J, Sommer T, BirchmeierW. (2002) Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nature of Cell Biology* 4(3): 222-31.

Goldstein B. (1992) Induction of gut in Caenorhabditis elegans embryos. *Nature.* 21;357(6375):255-7

Guo Y, Costa R, Ramsey H, Starnes T, Vance G, Robertson K, Kelley M, Reinbold R, Scholer H, Hromas R. (2002) The embryonic stem cell transcription factors Oct-4 and FoxD3 interact to regulate endodermal-specific promoter expression. *Proceeding of the National Academy of Sciences of the United States of America.* 99(6): 3663-7.

Harley, V.R., Lovell-Badge, R. and Goodfellow, P.N., (1994), Definition of a consensus DNA binding site for SRY, *Nucleic acid research.*, 22, 1500-1501

Heasman J, Wylie CC, Hausen P, Smith JC. (1984) Fates and states of determination of single vegetal pole blastomeres of X. laevis. *Cell.* 37(1):185-94

Henry, G.L. and Melton D.A., (1998) Mixer, a homeobox gene required for endoderm development. *Science* 281, 91-96

Hotta K, Takahashi H, Asakura T, Saitoh B, Takatori N, Satou Y, Satoh N. (2000) Characterization of Brachyury-downstream notochord genes in the Ciona intestinalis embryo. *Developmental Biology.* 224(1):69-80.

Hudson, C., Clements, D. Friday, R. V., Stott, D., Woodland, H. R., (1997) Xsox17a and b mediate endoderm formation in Xenopus. *Cell* 91,397-405

Ikeda, K., Shaw-White, J. R., Wert, S. E., and Whitsett, J. A. (1996). Hepatocyte nuclear factor 3 activates transcription of thyroid transcription factor 1 in respiratory epithelial cells. *Molecular Cell Biology.* 16, 3626-3636.

Imai K, Takada N, Satoh N, Satou Y. (2000) (beta)-catenin mediates the specification of endoderm cells in ascidian embryos. *Development.* 127 (14): 3009-20.

Joseph EM, Melton DA (1998) Mutant Vg1 ligands disrupt endoderm and mesoderm formation in Xenopus embryos. *Development.* 125 (14): 2677-85

Kanai Y, Kanai-Azuma M, Noce T, Saido TC, Shiroishi T, Hayashi Y, Yazaki K. (1996) Identification of two Sox17 messenger RNA isoforms, with and without the high mobility group box region, and their differential expression in mouse spermatogenesis. *Journal of Cell Biology.* 133(3):667-81.

Kanai-Azuma M, Kanai Y, Gad JM, Tajima Y, Taya C, Kurohmaru M, Sanai Y, Yonekawa H, Yazaki K, Tam PP, Hayashi Y. (2002) Depletion of definitive gut endoderm in Sox17-null

mutant mice. *Development* 129(10):2367-79

Kikuchi Y, Agathon A, Alexander J, Thisse C, Waldron S, Yelon D, Thisse B, Stainier DY. (2001) casanova encodes a novel Sox-related protein necessary and sufficient for early endoderm formation in zebrafish *Genes and Development*.15(12):1493-505.

Kikuchi Y., Trinh Le A., Reiter J.F., Alexander, J., Yelon. and Stainier, D.Y.R., (2000), The zebrafish bonnie and clyde gene encodes a Mix family homeodomain protein that regulates the generation of endodermal precursors, *Genes and Development* 14, 1279-1289

Kimelman D, Griffin KJ. (2000) Vertebrate mesendoderm induction and patterning. *Current Opinion on Genetic and Development*. 10(4):350-6

Kimmel CB, Warga RM, Schilling TF. 1990 Origin and organization of the zebrafish fate map; *Development*.;108(4):581-94.

Kobayashi K, Nishida H. (2001) Nuclear plasticity and timing mechanisms of the initiation of alkaline phosphatase expression in cytoplasm-transferred blastomeres of ascidians. *Developmental Biology*. 234(2): 510-20.

Kumano G, Nishida H. (1998) Maternal and zygotic expression of the endoderm-specific alkaline phosphatase gene in embryos of the ascidian, Halocynthia roretzi. *Developmental Biology* 198(2):245-52

Lin R, Hill RJ, Priess JR. (1998) POP-1 and anterior-posterior fate decisions in C. elegans embryos. *Cell* 92(2):229-39

Lin R, Thompson S, Priess JR (1995). pop-1 encodes an HMG box protein required for the specification of a mesoderm precursor in early C. elegans embryos. *Cell*. 83(4):599-609.

Maduro MF, Rothman JH. (2002) Making worm guts: the gene regulatory network of the Caenorhabditis elegans endoderm. *Developmental Biology.* 246(1):68-85. Review.

Mead, P. E., Zhou, Y., Lustig, K.D., Huber, T.L. and Kirschner, M.V. (1998) Cloning of mix-related homeodomain proteins using fast retrieval of gel shift activities, (FROGS), a technique for the isolation of dna-binding proteins. *Proceeding of the National Academy of Sciences of the United States of America.* 95, 11251-11256

Meneghini MD, Ishitani T, Carter JC, Hisamoto N, Ninomiya-Tsuji J, Thorpe CJ, Hamill DR, Matsumoto K, Bowerman B. (1999) MAP kinase and Wnt pathways converge to downregulate an HMG-domain repressor in Caenorhabditis elegans. *Nature.* 399 (6738):793-7.

Minoo P, Su G, Drum H, Bringas P, Kimura S. (1999) Defects in tracheoesophageal and lung morphogenesis in Nkx2.1(-/-) mouse embryos. *Developmental Biology.* 209(1):60-71.

Molkentin JD. (2000) The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression. *Journal Biological Chemistry.* 275(50): 38949-52. Review

Moreno-Mendoza N, Harley V, Merchant-Larios H., (2003) Cell aggregation precedes the onset of Sox9-expressing preSertoli cells in the genital ridge of mouse. *Cytogenetic Genome Research*;101(3-4):219-23.

Mori-Akiyama Y, Akiyama H, Rowitch DH, de Crombrugghe B. (2003) Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. *Proceeding of the National Academy of Sciences of the United States of Americ* U S A; 100(16):9360-5

Mushinski JF, Goodnight J, Rudikoff E, Morse HC 3rd, Langdon WY. (1994) Expression of c-cbl proto-oncogene is modulated during differentiation but not during induction of proliferation. *Oncogene.* 9(9): 2489-97

Nieuwkoop PD. (1973) The organization center of the amphibian embryo: its origin, spatial organization, and morphogenetic action. *Advances in Morphology*;10:1-39

Nishida H. (1996) Vegetal egg cytoplasm promotes gastrulation and is responsible for specification of vegetal blastomeres in embryos of the ascidian Halocynthia roretzi. *Development.* 122(4):1271-9

Nishida H. (1997) Cell fate specification by localized cytoplasmic determinants and cell interactions in ascidian embryos. *International Review of Cytology.*176:245-306. Review

Nishida, H. (1987) Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to tissue-restricted stage. *Developmental Biology.* 121, 526-541

Nishida, H. (1992) Developmental potential for tissue differentiation of fully dissociated cells of the ascidian embryo. *Roux's Arch.Dev.Biol.* 210, 81-87

Nishida, H. (1993) Localized regions of egg cytoplasm that promote expression of endoderm-specific alkaline phosphatase in embryos of the ascidian Halocynthia roretzi. *Development* 118 : 1-7.

Nishida, H., and Satoh, N. (1983) Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. I. Up to the eight-cell stage. *Developmental Biology* 99, 382-394

Nishida, H., and Satoh, N. (1985) Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. II. The 16- and 32-cell stage. *Developmental Biology* 110, 440-454

Osada SI, Wright CV (1999) Xenopus nodal-related signaling is essential for mesendodermal patterning during early embryogenesis. *Developmen*t. 126(14): 3229-40.

Patient RK, McGhee JD (2002) The GATA family (vertebrates and invertebrates). *Current Opinion of Genetic and Development.* 12(4): 416-22. Review.

Pearce JJ, Evans MJ. (1999) Mml, a mouse Mix-like gene expressed in the primitive streak. *Mechanism of Development* 87(1-2): 189-92.

Reim G, Mizoguchi T, Stainier DY, Kikuchi Y, Brand M. (2004) The POU domain protein spg (pou2/Oct4) is essential for endoderm formation in cooperation with the HMG domain protein casanova. *Developmental Cell.* 6(1): 91-101

Reiter JF, Kikuchi Y, Stainier DY. (2001) Multiple roles for Gata5 in zebrafish endoderm formation. *Development.* 128(1): 125-35

Ristoratore F., Spagnolo A., Aniello, F., Branno, M., Fabrini, F. and Di Lauro R., (1999) Expression and Functional analysis of Cititf1, an ascidian NK2 class gene, suggest its role in endoderm development *Development.* 126 (22): 5149-59

Rocheleau CE, Yasuda J, Shin TH, Lin R, Sawa H, Okano H, Priess JR, Davis RJ, Mello CC. (1999) WRM-1 activates the LIT-1 protein kinase to transduce anterior/posterior polarity signals in C. elegans. *Cell.* 97(6): 717-26.

Ruffins SW, Ettensohn CA. (1996) A fate map of the vegetal plate of the sea urchin (Lytechinus variegatus) mesenchyme blastula. *Development* 122(1):253-63

Russ AP, Wattler S, Colledge WH, Aparicio SA, Carlton MB, Pearce JJ, Barton SC, Surani MA, Ryan K, Nehls MC, Wilson V, Evans MJ (2000) Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature.* 404 (6773): 95-9

Satou Y, Imai KS, Satoh N. (2001) Early embryonic expression of a LIM-homeobox gene Cs-lhx3 is downstream of beta-catenin and responsible for the endoderm differentiation in Ciona savignyi embryos. *Development.* 128(18): 3559-70

Schier AF, Neuhauss SC, Helde KA, Talbot WS, Driever W. (1997) The one-eyed pinhead gene functions in mesoderm and endoderm formation in zebrafish and interacts with no tail. *Development.* 124(2):327-42.

Shen JH, Ingraham HA. (2002) Regulation of the orphan nuclear receptor steroidogenic factor 1 by Sox proteins.*Molecular Endocrinology* ;16(3):529-40;

Sherwood DR, McClay DR. (2001) LvNotch signaling plays a dual role in regulating the position of the ectoderm-endoderm boundary in the sea urchin embryo. *Development.* 128 (12): 2221-32.

Simmen MW, Leitgeb S, Clark VH, Jones SJ, Bird A. (1998) Gene number in an invertebrate chordate, Ciona intestinalis. *Proceeding of the National Academy of Sciences of the United States of America.* 95(8): 4437-40

Spagnuolo A, Di Lauro R. (2002) Cititf1 and endoderm differentiation in Ciona intestinalis. *Gene.* 287 (1-2): 115-9.

Tada M, Casey ES, Fairclough L, Smith JC. (1998) Bix1, a direct target of Xenopus T-box genes, causes formation of ventral mesoderm and endoderm. *Development.* 125(20): 3997-4006

Takada S, Koopman P. (2003). Origin and possible roles of the Sox8 transcription factor gene during sexual development. *Cytogenetic Genome Research* 101(3-4):212-8. Review

Tomioka M, Nishimoto M, Miyagi S, Katayanagi T, Fukui N, Niwa H, Muramatsu M, Okuda A.(2002) Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. *Nucleic Acids Research.* 30(14):3202-13

Warga RM, Nusslein-Volhard C. (1999) Origin and development of the zebrafish endoderm. *Development.* 126(4): 827-38

Weber H, Symes CE, Walmsley ME, Rodaway AR, Patient RK. (2000) A role for GATA5 in Xenopus endoderm specification. *Development.* 127 (20): 4345-60

Weeks DL, Melton DA (1987) A maternal mRNA localized to the vegetal hemisphere in Xenopus eggs codes for a growth factor related to TGF-beta. *Cell.* 51(5): 861-7.

Wegner M. From head to toes: the multiple facets of Sox proteins. (1999) *Nucleic Acids Research* 27(6): 1409-20. Review

Whittaker JR. (1977) Segregation during cleavage of a factor determining endodermal alkaline phosphatase development in ascidian embryos. *Journal of Experimental Zoology* 202(2):139-53.

Whittaker, J.R. (1973) Segregation during ascidian embryogenesis of egg cytoplasmic information for tissue specific enzyme development. *Proceeding of the National Academy of Sciences of the United States of America* 70: 2096-2100

Xanthos, J.B., Kofron, M., Wylie, C. and Heasman J., (2001) Maternal vegT is the initiator of a molecular network specifying endoderm in Xenopus laevis. *Development*, 128, 167-180

Yamada L, Kobayashi K, Degnan B, Satoh N, Satou Y. (2003) A genomewide survey of developmentally relevant genes in Ciona intestinalis. IV. Genes for HMG transcriptional regulators, bZip and GATA/Gli/Zic/Snail. *Development Genes and Evolution* 213 (5-6): 245-53.

Yasuo H, Lemaire P. (1999) A two-step model for the fate determination of presumptive endodermal blastomeres in Xenopus embryos; *Current Biology* 9(16): 869-79.

Yeo C, Whitman M. (2001) Nodal signals to Smads through Cripto-dependent and Cripto-independent mechanisms. *Molecular Cell.* 7(5): 949-57.

Yokouchi M, Kondo T, Houghton A, Bartkiewicz M, Horne WC, Zhang H, Yoshimura A, Baron R. (1999) Ligand-induced ubiquitination of the epidermal growth factor receptor involves the interaction of the c-Cbl RING finger and UbcH7. *Journal of Biological Chemistry.* 274(44): 31707-12.

Zhang J, Houston DW, King ML, Payne C, Wylie C, Heasman J. (1998) The role of maternal VegT in establishing the primary germ layers in Xenopus embryos. *Cell* 94 (4): 515-24.

Zhang J, Talbot WS, Schier AF. (1998) Positional cloning identifies zebrafish one-eyed pinhead as a permissive EGF-related ligand required during gastrulation. *Cell.* 92(2): 241-51

Zorn, A.M., Barish, G.D., Williams, B.O., Lavender, P., Klymkowsky, M.W., and Varmus, H.E., (1999) Regulation of Wnt signaling by sox proteins: Xsox17a/b and Xsox3 physically interact with β-catenin, *Molecular Cell* 4, 487-498

Acknowledgements

During the time I have spent at the SZN I recived help from several persons to whom I am grateful: the staff of the SZN molecular biology service, headed by Dr. Elio Biffali, who provided a fast and efficient sequence service and synthesised all of the oligonucleotides used in this study, particularly the technicians Elvira Mauriello and Fabrizio Campili; the staff of in situ service headed to Dr. Rita Marino. I should also like to thank the technician of my own lab Michele Pischetola for his tecnical support. Special thanks go to Rosanna Parlato and Franceso Errico for continuos advice, support and friendship during my stay to SZN. Thanks also to my SZN supervisor Dr. Antonietta Spagnuolo and Prof. Dennis Duboule, and particularly my director of studies Prof. Roberto Di Lauro for fruitful scientific discussion and his advice.